



Universitat de Girona
Escola Politècnica Superior

Projecte/Treball Final de Carrera

Estudi: Eng. Tècn. Informàtica de Gestió. Pla 2001

Títol: Motor de cerca iSAC (Servei intel·ligent d'atenció ciutadana via web)

Document: Memòria

Alumne: Albert Sabrià i Torrent

Director/Tutor: Miquel Montaner Rigall
Departament: Extern a l'EPS
Àrea:

Convocatòria (mes/any): 09/2006

Contingut

PART I.....	7
Descripció, origen i objectius del projecte	7
Capítol 1. Introducció	8
1.1. Introducció del sistema	8
1.2. Els beneficis	10
1.3. Els col·laboradors	11
Capítol 2. El sistema a implantar	12
2.1. Objectius	12
PART II.....	15
Desenvolupament de sistemes d'informació.....	15
Capítol 1. Metodologia de planificació i desenvolupament	16
1.1. La mètrica versió 3	16
1.2. Metodologia utilitzada	18
Capítol 2. Estudi de viabilitat del sistema (EVS).....	19
2.1. Establir l'abast del sistema	19
2.1.1. Estudi de la sol·licitud	19
2.1.2. Definició del pla de treball.....	20
2.2. Estudi de la situació actual	22
2.2.1. Valoració de l'estudi de la situació actual	22
2.3. Definició de requisits del sistema	22
2.4. Estudi d'alternatives per la solució	24
2.5. Selecció de la solució	28
Capítol 3. Anàlisi del sistema d'informació (ASI)	29
3.1. Definició del sistema.....	29
3.1.1. Determinar l'abast del sistema.....	29
Diagrama de context del sistema	31
3.1.2. Procés de resposta en l'atenció automatitzada.....	32
3.1.3. El motor de cerca iSAC	32
3.2. Establiment de requisits	43
3.2.1. Obtenció de requisits	43
3.2.2. Especificació de casos d'ús.....	46
3.3. Anàlisi dels casos d'ús	48
3.3.1. Identificació de classes associades a un cas d'ús.....	48

3.4.	Anàlisi de classes	49
3.5.	Elaboració del model de dades.....	51
3.5.1.	Elaboració del model lògic de dades	51
	Diagrama Entitat / Relació.....	51
	Model relacional	52
3.6.	Definició d'interfícies d'usuari	54
3.6.1.	Especificació de principis generals de la interfície.....	54
3.6.2.	Especificació de formats individuals de la interfície de pantalla.....	54
3.6.3.	Identificació de perfils i diàlegs.....	56
Capítol 4.	Disseny del sistema d'informació (DSI)	57
4.1.	Definició de l'arquitectura del sistema	57
4.2.	Disseny de casos d'ús reals	62
4.3.	Disseny de classes	63
4.4.	Disseny físic de dades	64
4.4.1.	Transformació del model de classes al model físic de dades.....	64
4.5.	Especificacions de construcció	66
4.5.1.	Especificacions de l'entorn de construcció.....	66
4.5.2.	Llenguatge de programació.....	66
4.5.3.	Entorn de desenvolupament	67
4.5.4.	Navegador	68
	Instal·lació barra d'eines de Microsoft Internet Explorer	69
	Instal·lació barra d'eines de Mozilla Firefox	71
4.5.5.	Altres eines software.....	73
	Eines de disseny:.....	73
	Transferència de fitxers al servidor:	74
Capítol 5.	Construcció del sistema d'informació (CSI)	75
5.1.	Preparació de l'entorn de generació i construcció.	75
5.2.	Execució de les proves	75
5.3.	Elaboració dels manuals d'usuari.....	76
5.4.	Definició de la formació de l'usuari final	76
PART III	77	
Ampliacions i millores	77	
Capítol 1. Ampliacions i millores	78	
PART IV.....	80	
Conclusions.....	80	
Capítol 1. Conclusions.....	81	

PART V	82
Documentació de referència.....	82
Capítol 1. Bibliografia	83
Capítol 2. Referències web.....	84

Índex de figures

FIGURA 1: PLANNING DEL PROJECTE	21
FIGURA 2: ALTERNATIVA DE SOLUCIÓ: “LA CAIXA”	24
FIGURA 3: ALTERNATIVA DE SOLUCIÓ: TELEFÓNICA ESPANYA	25
FIGURA 4: ALTERNATIVA DE SOLUCIÓ: SARA	27
FIGURA 5: DIAGRAMA DE CONTEXT DEL SISTEMA	31
FIGURA 6: PANTALLA D’INICI.....	34
FIGURA 7: EXEMPLE DE CORRECCIÓ ORTOGRÀFICA	36
FIGURA 8: RESULTATS DEL SISTEMA HAVENT LEMMATITZAT	38
FIGURA 9: GRÀFICA CORRESPONENT A LA FÓRMULA DEL CÀLCUL D’AFINITAT	39
FIGURA 10: ESQUEMA DEL REFINAMENT	42
FIGURA 11: DIAGRAMA DE CASOS D’ÚS	45
FIGURA 12: DIAGRAMA DE CLASSES DE LES FAQs	49
FIGURA 13: DIAGRAMA DE CLASSES DELS DICCIONARIS	50
FIGURA 14: DIAGRAMA ENTITAT / RELACIÓ	51
FIGURA 15: CAPÇALERA DEL SISTEMA iSAC	54
FIGURA 16: PANTALLA D’INCI.....	54
FIGURA 17: PANTALLA DE LLISTAT DE RESULTATS	55
FIGURA 18: PANTALLA AMB EL RESULTAT	56
FIGURA 19: DIAGRAMA DE PERFILS I DIÀLEGS	56
FIGURA 20: ARQUITECTURA iSAC	58
FIGURA 21: MÒDULS DE L’ARQUITECTURA iSAC	59
FIGURA 22: ARQUITECTURA iSAC PER A UN AJUNTAMENT GRAN	60
FIGURA 23: ARQUITECTURA iSAC (VERSIÓ EXTENSA) AJUNTAMENT PETIT	61
FIGURA 24: DIAGRAMA DE CASOS D’ÚS	62
FIGURA 25: DIAGRAMA DE CLASSES DELS DICCIONARIS	63
FIGURA 26: DIAGRAMA DE CLASSES DE LES FAQs	63
FIGURA 27: BARRA D’EINES iSAC	69

Introducció

Aquest document és la memòria del projecte fi de carrera titulat Motor de cerca iSAC (servei intel·ligent d'atenció ciutadana via web) realitzat per Albert Sabrià i Torrent, estudiant d'Enginyeria Tècnica d'Informàtica de Gestió de la Universitat de Girona.

La realització d'aquest document ha estat feta de manera que el lector pugui arribar a tenir coneixença amb detall de totes les tasques realitzades en el desenvolupament del projecte així com les tècniques i metodologies emprades en ell.

Part I

Descripció, origen i objectius del projecte

En aquesta part s'exposa una introducció sobre l'origen del projecte i també els objectius que vol complir.

Capítol 1. Introducció

En aquest capítol es desglossen els següents apartats:

- Introducció del sistema
- Els beneficis
- Els col·laboradors

1.1. Introducció del sistema

El projecte iSAC (Servei Intel·ligent d'Atenció Ciutadana via web) es va iniciar el mes de gener de 2006 amb l'ajut del nou coneixement científic en agents intel·ligents, junt amb l'aplicació de les Tecnologies de la Informació i la Comunicació (TIC) i els cercadors.



Actualment, el servei actual d'atenció al ciutadà està compostat per dues àrees:

- L'atenció directa a les oficines
- L'atenció telefònica a través del Call Center

Les limitacions de personal i horari d'atenció fan que aquest servei perdi eficàcia.

El servei d'atenció ciutadana està saturat i seguirà saturat. Els funcionaris no tenen temps de fer res més que respondre al telèfon, i es senten absolutament condicionats a reaccionar davant de la demanda elevada, creixent i canviant qualitativament, quan el servei es defineix com a proactiu i orientat a les persones. Volen respondre-li de forma ràpida, precisa i empàtica, i ell s'expressa a la seva manera, en el seu propi idioma i d'acord amb la seva cultura.

Es tracta de resoldre el problema de saturació dels serveis d'atenció al ciutadà amb l'automatització de la resposta a les demandes més habituals i previsibles, per permetre alliberar personal cap a feines de back office, coordinació, avaluació de serveis, lectura del feedback ciutadà, i fer possible una millora continuada dels serveis públics i una resposta adequada als temps i a la qualitat de vida actual.

Es vol desenvolupar un producte amb una tecnologia capaç d'ampliar i millorar la capacitat i la qualitat de l'atenció ciutadana en les administracions públiques, sigui quina sigui la seva dimensió. Tot i això, aquest projecte l'explotaran especialment els ajuntaments, als quals la ciutadania s'acosta amb tot tipus de preguntes i dubtes, habitualment no restringides a l'àmbit local.

Alguns exemples podrien ser:

- Quin és el recorregut de la cavalcada de reis?
- Quin dia canvien l'hora?
- Quan haig de matricular el meu fill a l'escola?
- Quina és l'adreça del centre d'activitats teatrals?
- ...

Més concretament, es vol automatitzar a través d'un portal web l'atenció al ciutadà per tal d'obtenir un servei més efectiu.

El nou servei web vol trencar les limitacions esmentades mentre que es millora també la informació proveïda i la personalització d'aquesta a les necessitats de cada ciutadà. Com a objectiu més ambiciós del projecte destaca la humanització del servei, és a dir, que el ciutadà se senti atès com si una persona estigués interactuant amb ell i responent a les seves qüestions.

Així doncs, aquest projecte pretén desplegar un software lliure de tecnologia de buscadors de FAQ (en anglès, Frequent Asked Questions) per oferir-lo a una ciutadania que sol·licita diàriament aquest servei. Aquestes FAQ es crearan des de l'experiència existent als serveis municipals d'Atenció Ciutadana.

D'impacte immediat en l'atenció directa efectiva, la millora de la qual serà percebuda i valorada per la ciutadania, per la reducció dels temps d'espera i per la millora de l'eficàcia en la resposta adequada.

El concepte ciutadania que ens referim en el projecte és el més ampli possible, i inclou de manera natural a tots els que utilitzen els serveis d'atenció, i per tant també als usuaris interns de l'Ajuntament, que l'utilitzaran per consultar o demanar ajut, ja sigui quan tenen problemes estrictament tècnics, logístics, de tramitació, o d'altra mena que es pugui preveure.

En aquesta primera fase, el projecte es realitzarà especialment per les administracions de Catalunya i per tant en idioma català.

Està clar doncs que s'intenta oferir un servei a una ciutadania extensa i molt diferent entre ells. S'ha de crear un sistema capaç d'entendre a tothom, s'expressi com s'expressi, tot i que utilitzi paraules diferents a les que nosaltres podríem utilitzar. L'ambició és que dos usuaris demanin el mateix, utilitzant paraules diferents, i que el sistema sigui capaç de comprendre'ls i retornar el resultat desitjat.

1.2. Els beneficis

Els beneficis mesurables del marc iSAC són els següents:

- Oferirà una resposta preparada amb anterioritat per un expert, i per tant, la millor possible.
- Respondrà a preguntes inter administració que transcendeixin l'àmbit local i que per tant, són de difícil resposta a l'actualitat.
- Previndrà de demandes estacionals ocasionals. Oportunitat d'avançar informació.
- Facilitarà la alliberació de personal del front office per l'atenció automatitzada de les peticions habituals i repetitives.
- Facilitarà el dedicar més personal al back office (avaluació i gestió per assegurar el bon funcionament del front office, l'actualització del servei, la innovació i una major proactivitat).
- Facilitarà la derivació de més personal cap a les demandes complexes i a l'atenció de col·lectius amb necessitats especials.
- Reduirà el temps de formació del nou personal del servei, per l'experiència de la base de coneixement, i permetrà la seva incorporació més senzilla i ràpida a l'atenció personal.
- Ajudarà a definir millor els models d'atenció ciutadana, i crear estàndards que afavoreixin la seva evolució i sobretot la seva expansió als ajuntaments que encara no en disposin.
- Impulsarà la creació d'eines complementaries, estandarditzades i modulars, basades en software lliure i no propietària per ser implementades a qualsevol ajuntament o organització pública.
- Permetrà la creació de mòduls de formació col·lectiva i de reflexió sobre el model propi per ajudar als ajuntaments en la creació i la millora dels serveis d'atenció ciutadana.
- Ampliació de l'horari de servei a 24 h en les demandes més freqüents
- Augment de la satisfacció de la ciutadania
- Reducció de costos d'atenció directa
- Millora de l'entorn de treball, reducció de pressió

Pel que fa a l'usuari:

- Comprendre'l quan parla un llenguatge col·loquial, o no utilitza les paraules correctes per no conèixer l'argot de l'administració pública.
- Identificar la seva demanda encara que tingui errors d'ortografia.
- Atendre al ciutadà en qualsevol demanda, encara que formuli preguntes que no transcendeixin a la responsabilitat de l'administració pública.

1.3. Els col·laboradors

Per tal de desenvolupar aquest projecte ha estat necessària la col·laboració estreta amb l'Ajuntament de Terrassa.



L'Ajuntament de Terrassa ha aportat tot el coneixement específic sobre l'Administració Pública, personal dedicat a organitzar, reestructurar i millorar les FAQs, el mètode de treball del servei i un entorn de proves vàlid per tal de testejar el projecte.

Capítol 2. El sistema a implantar

En aquest capítol es defineixen els objectius principals que ha de contemplar el projecte a realitzar i les seves característiques generals.

2.1. Objectius

L'objectiu del projecte és el desenvolupament del motor de cerca del sistema iSAC.

Quan una persona, hagi de realitzar una consulta a l'atenció ciutadana, no serà un impediment el dia i l'hora que sigui. A través del portal web d'atenció ciutadana, qualsevol persona podrà realitzar la consulta que vulgui, en el moment que desitgi.

Per fer-ho, només haurà d'accedir al portal i introduir la consulta desitjada, i un cop retornats els resultats per part del sistema, escollir la més adient pels seus interessos.

Però aquest procés simple per l'usuari, es torna complex per part del sistema. La consulta introduïda per l'usuari no es tracta directament, sinó que ha de passar per un seguit de filtres i tractaments que faran que la resposta s'ajusti el màxim a "què vol el ciutadà?".

Amb això ens referim, a que tothom s'expressa de la seva manera, utilitzant el seu llenguatge col·loquial, i fins i tot amb paraules típiques i pròpies de la seva regió. També es poden produir errors ortogràfics, gramaticals, semàntics,... i això no ha de ser un impediment pel sistema.

Mitjançant tècniques de llenguatge natural, el sistema ha de ser capaç de proposar i corregir aquests errors, extreure tot tipus de paraules que no aportaran informació (articles, determinants,...), obtenir el lema de les paraules per així englobar totes les seves diferents formes, funcions i possibles temps verbals, obtenir els sinònims i el significat de les expressions pròpies de la llengua, així com expressions que utilitzem normalment amb llenguatge verbal com poden ser "demà", "el dia de reis",...

Però les coses no s'acaben amb aquest pretractament, ja que mitjançant el feedback que es produeix per part de l'usuari, el sistema aprèn. A mesura que s'utilitza el sistema, es reconsideren les importàncies de les FAQs segons el número de consultes i per tant es tindran més en compte aquelles consultes més freqüents i importants.

Amb tot això, el sistema mostrarà a l'usuari, un llistat de FAQs que s'ajusten a la demanda realitzada, mostrant-les de manera ordenada segons l'afinitat. El càlcul d'aquesta afinitat serà un procés ampli i segurament variant a mida que va avançant el projecte. Es calcularà segons el nombre de paraules coincidents, la importància d'aquestes paraules dins l'estructura, la importància també de les respostes, del context de la consulta,...

Altres aspectes com podrien ser la gestió dels diferents diccionaris i FAQs, així com la integració dins l'estructura de bases de dades que estan en funcionament actualment en les oficines d'atenció ciutadana, queden fora de l'abast del projecte.

En resum, podríem dir que els objectius del projecte serien:

- Desenvolupar el motor de cerca del sistema iSAC.
- Identificar la pregunta i la resposta als repositoris de FAQs prèviament definits pel Servei d'Atenció Ciutadana.
- Entendre correctament la pregunta que se li formula. Això comporta:
 - Entendre el llenguatge col·loquial.
 - Preveure errors d'ortografia.
 - Poder millorar la cerca mitjançant:
 - Sinònims.
 - Expressions populars.
 - Expressions de temps.
 - Regionalismes.
 - Extreure el contingut important de la cerca.
 - Englobar totes les possibles conjugacions i formes que pot tenir cada paraula.
- Aprenentatge del sistema per millorar en les següents cerques.
- Ésser utilitzat per tots els ciutadans.
- Respondre al ciutadà amb la resposta encertada.

A més, un altre objectiu d'aquest projecte, no menys important que la resta, consisteix en la necessitat de generar codi open source o codi lliure respecte a una tecnologia, la de cercadors i recomanadors, que es sent fortament protegida per productes comercials estrangers, inaccessible per a la gran majoria d'administracions públiques locals. El gran interès del projecte és fer accessible els esmentats desenvolupaments en benefici dels ciutadans que tenen en l'administració pública local la seva primera finestra a l'administració.

També ho serà per tant el fet d'aprofitar-se d'altres aplicacions que ens poden ajudar a reduir temps alhora de desenvolupar el nostre sistema. L'exemple més clar, és la utilització del diccionari català de OpenOffice. Aquest mòdul OpenSource ens ajudarà a simplificar molt les possibles correccions ortogràfiques que l'usuari pugui haver introduït a la seva consulta. També s'ha utilitzat el diccionari FreeLing per tal d'obtenir els lexemes de les paraules.

Amb tot això remarcar, que es desitja desenvolupar i posar en funcionament l'aplicació utilitzant eines gratuïtes per tal que no suposi un cost afegit a l'administració pública posar en funcionament aquest sistema.

Part II

Desenvolupament de sistemes d'informació

En aquesta part del document s'exposa l'aplicació d'una metodologia d'anàlisi i disseny de sistemes d'informació en el projecte actual: la Mètrica v.3.

Capítol 1. Metodologia de planificació i desenvolupament

En aquest capítol es descriu la metodologia mètrica versió 3 i es defineix la metodologia utilitzada per realitzar el procés d'informació del sistema.

1.1. La mètrica versió 3

Aquesta és una metodologia desenvolupada pel Ministeri d'Administracions Públiques.



Els principals objectius d'aquesta metodologia són:

- Proporcionar o definir Sistemes d'Informació que ajuden a aconseguir els fins de l'Organització mitjançant la definició d'un marc estratègic pel desenvolupament dels mateixos.
- Dota a l'Organització de productes software que satisfacin les necessitats dels usuaris donant una major importància a l'anàlisi de requisits.
- Millorar la productivitat dels departaments de Sistemes i Tecnologies de la Informació i les Comunicacions, permeten millorar la capacitat d'acceptació als canvis i tenint en compte la reutilització.
- Facilitar la comunicació i enteniment entre els diferents participants en la producció de software al llarg del cicle de vida del projecte, tenint en compte el seu parer i responsabilitat, així com les necessitats de tots i cada un d'ells.
- Facilitar l'operació, manteniment i ús dels productes software obtinguts.

La metodologia descompon cada un dels processos en activitats, i a la vegada en feines. Per cada feina es descriu el seu contingut fent referència a les seves principals accions, productes, tècniques, pràctiques i participants.

Els processos de l'estructura principal de MÈTRICA Versió 3 són els següents:

- PLANIFICACIÓ DE SISTEMES D'INFORMACIÓ

L'objectiu d'aquest procés és proporcionar un marc estratègic de referència pels sistemes d'Informació d'un determinat àmbit de l'organització.

- **DESENVOLUPAMENT DE SISTEMES D'INFORMACIÓ**

Conté totes les activitats i feines que s'han de portar a terme per desenvolupar un sistema, cobrint des de l'anàlisi de requisits fins la instal·lació del software.

Per facilitar la comprensió i donada la seva amplitud i complexitat se subdivideix en cinc processos:

- **ESTUDI DE VIABILITAT DEL SISTEMA**

El seu propòsit és analitzar un conjunt concret de necessitats, amb la idea de proposar una solució a curt termini. Els criteris amb els que es fa aquesta proposta no seran estratègics sinó tàctics i relacionats amb aspectes econòmics, tècnics, legals i operatius.

- **ANÀLISIS DEL SISTEMA D'INFORMACIÓ**

El propòsit d'aquest procés és aconseguir l'especificació detallada del sistema d'informació, a través d'un catàleg de requisits i una sèrie de models que cobreixin les necessitats d'informació dels usuaris pels que es desenvoluparà el sistema d'informació i que serà l'entrada pel següent procés.

- **DISSENY DEL SISTEMA D'INFORMACIÓ**

El seu propòsit és obtenir la definició de l'arquitectura del sistema i de l'entorn tecnològic que li donarà suport, junt amb l'especificació detallada dels components del sistema d'informació.

- **CONSTRUCCIÓ DEL SISTEMA D'INFORMACIÓ**

Aquest procés té com objectiu final la construcció i prova dels diferents components del sistema d'informació, a partir del conjunt d'especificacions lògiques i físiques del mateix. Es desenvolupen els procediments d'operació i seguretat i elaboració dels manuals de l'usuari final i d'exploació.

- **IMPLANTACIÓ I ACCEPTACIÓ DEL SISTEMA**

Aquest procés té com objectiu principal, el lliurament i acceptació del sistema en la seva totalitat i un segon objectiu que és portar a terme les activitats oportunes pel pas a producció del sistema.

- **MANTENIMENT DE SISTEMES D'INFORMACIÓ.**

L'objectiu d'aquest procés és l'obtenció d'una nova versió d'un sistema d'informació desenvolupada amb MÈTRICA, a partir de les peticions de manteniment que els usuaris realitzin amb motiu d'un problema detectat en el sistema o per la necessitat d'una millora del mateix.

1.2. Metodologia utilitzada

Per la planificació i desenvolupament del projecte, el primer plantejament va ser utilitzar la metodologia MÈTRICA versió 3.

Aquesta va ser la primera impressió ja que la necessitat del Pla de Sistemes d'Informació es basa en el fet de tenir un projecte ampli, de constant millora i ampliacions, i el fet que en aquestes diferents etapes, col·labori i participi gent diferent.

Aquest pla ens hauria permès establir unes pautes i una base en la qual recolzar tot el projecte en el present i el futur. Un projecte innovador com aquest hauria de disposar del Pla de Sistemes d'Informació.

A mida que s'avançava amb el projecte, me'n vaig adonar que aquesta metodologia és massa complexa per desenvolupar una petita part d'un projecte com aquest. El problema amb el qual em vaig trobar al implantar aquesta metodologia recau sobre el fet que aquesta està bàsicament pensada per a projectes "sencers" i per organitzacions "grans" on intervenen diferents departaments/treballadors. És per això que en el desenvolupament d'aquest projecte s'han seguit alguns apartats i línies de treball de la metodologia, i s'han obviat altres apartats per considerar-los menys importants o inapropiats pel tipus de projecte que és.

Així doncs, la conclusió ha estat utilitzar la base de la metodologia mètrica versió 3 aprovada pel Ministeri d'Administracions públiques, tot i que obviant algun dels seus apartats.

Capítol 2. Estudi de viabilitat del sistema (EVS)

En aquest capítol s'exposa l'estudi de viabilitat del sistema, segons el que indica l'apartat de Mètrica v.3 que té el mateix nom. En aquest apartat, la Mètrica v.3 proposa estudiar les següents activitats:

- Establir l'abast del sistema.
- Estudi de la situació actual.
- Definició de requisits del sistema.
- Estudi d'alternatives de solució.
- Valoració de les alternatives.
- Selecció de la solució.

2.1. Establir l'abast del sistema

2.1.1. Estudi de la sol·licitud

Es vol implementar i posar en marxa una aplicació web basada en software lliure de tecnologia de buscadors de FAQ per oferir-lo a la ciutadania, que sol·licita diàriament aquest servei.

L'abast del projecte és extens i sempre estarà pendent de millores i innovacions per tal de no quedar obsolet dins l'evolució de les noves tecnologies de la informació.

Per acotar el projecte, hem desglossat aquest abast, i ens centrarem principalment en el motor de cerca.

El sistema ha de ser capaç de donada una cadena d'entrada, donar la resposta més adient segons les paraules introduïdes a aquesta cadena.

Aquest procés consisteix en un seguit de tècniques de llenguatge natural, que hauran de sobreposar-se a possibles errors ortogràfics, així com intentar extreure què es demana, obviant algunes paraules i transformant i afegint-ne d'altres.

Aquest procés ha de finalitzar amb una interacció del sistema amb l'usuari, per tal que amb el pas del temps, el sistema pugui aprendre de possibles errors, i així millorar cada vegada més la cerca dins del repositori.

2.1.2. Definició del pla de treball

Per simplificar les coses, hem desglossat el motor de cerca en diferents fases.

- Fase 1: Disseny de la base de dades
 - Model entitat / relació
- Fase 2: Correcció ortogràfica
 - Eines i diccionaris OpenSource
 - Disseny de la base de dades ortogràfica
 - Integració de les eines OpenSource
- Fase 3: Filtratge, Completament i Lemmatització de la pregunta
 - StopWords
 - Classificació de paraules
 - Sinònims
 - Expressions i frases populars
 - Lemmatització
- Fase 4: Algoritme de cerca
 - Algorisme de càlcul de rànquings
 - Presentació de la llista de resultats

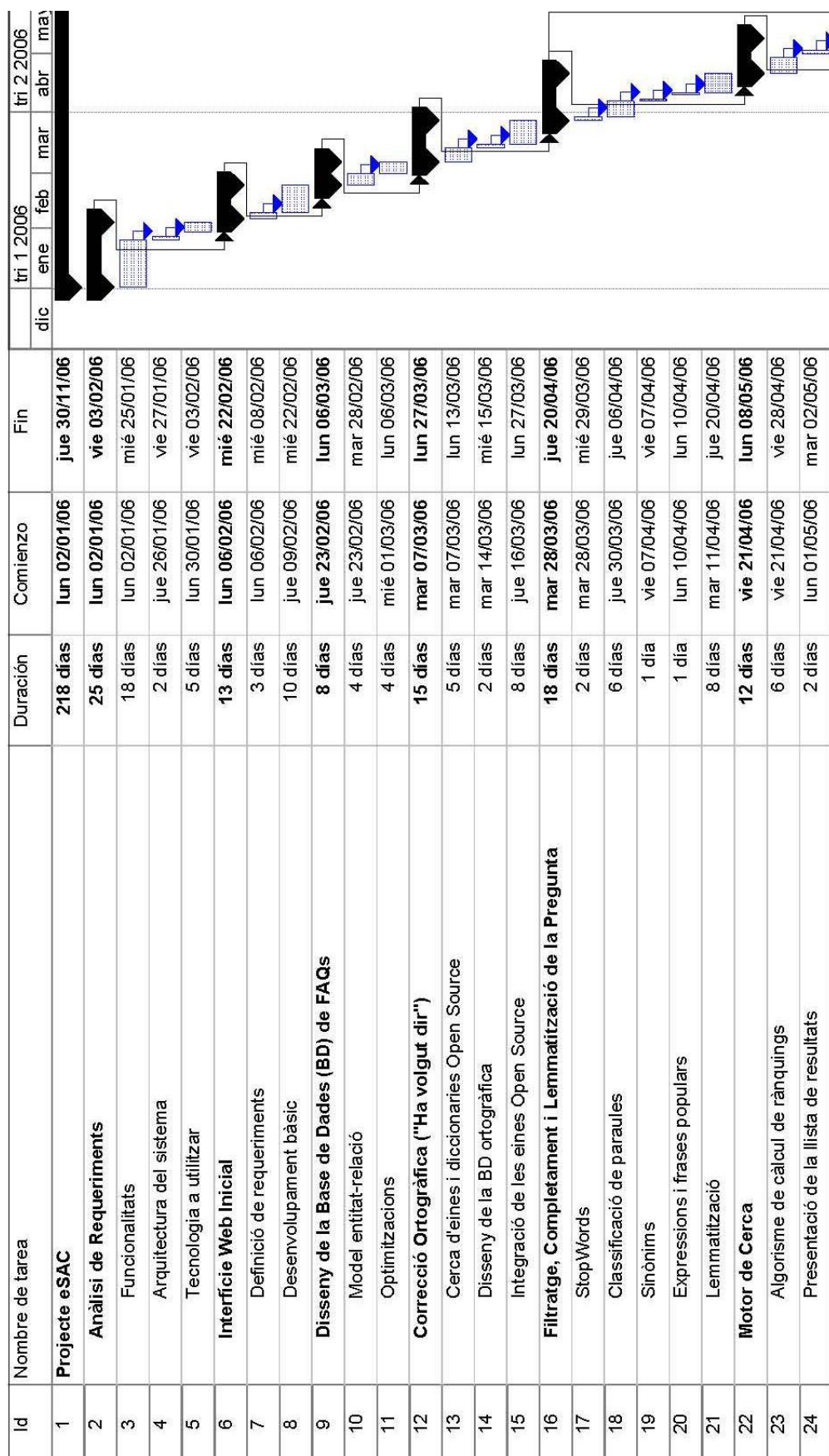


Figura 1: Planning del projecte

2.2. Estudi de la situació actual

2.2.1. Valoració de l'estudi de la situació actual

Actualment les administracions públiques no disposen d'un software o aplicació, ja sigui via web o no, per tal que el ciutadà pugui satisfer les seves qüestions.

El servei actual d'atenció al ciutadà està compostat per dues àrees:

- Atenció telefònica a través del Call Center
- Atenció directa a les oficines

Tot ciutadà pot adreçar-se a la corresponent administració pública a través de la via telefònica, atenent-se a les conseqüències que això comporta com pot ser la saturació de la línia telefònica o la manca de personal per atendre les trucades.

Una altra possible solució que actualment existeix, és presentar-se físicament a l'oficina d'atenció ciutadana. Això comporta un alt handicap com és haver-se de desplaçar, però sobretot el fet de trobar-se un elevat volum de ciutadans que també necessiten l'atenció ciutadana per obtenir tràmits i d'altres aspectes que solament es poden aconseguir personificant-se a l'oficina i per tant haver-se d'atenir a un temps d'espera totalment no desitjat.

A tot això, s'ha d'afegir el fet de les limitacions d'horari, i el possible maldecap que això comporta per aquelles persones amb horari laboral similar.

Totes aquestes limitacions fan que aquest servei perdi eficàcia.

2.3. Definició de requisits del sistema

El nou servei web vol trencar aquestes limitacions mentre que es millora també la informació proveïda i la personalització d'aquesta a les necessitats de cada ciutadà. Com a objectiu més ambiciós del projecte destaca la humanització del servei, és a dir, que el ciutadà se senti atès com si una persona estigués interactuant amb ell i responent a les seves qüestions.

El problema identificat en els serveis d'atenció ciutadana de l'administració local es centra en la voluntat d'atendre millor les persones que s'hi adrecen en un context d'incapacitat per ampliar el personal i de creixement constant i ràpid de la demanda més eficient i automatitzada.

La solució que la nova tecnologia ofereix és respondre les demandes les 24 hores del dia des de qualsevol indret de món, disposant solament d'un punt de connexió a la xarxa de les xarxes: internet; a comoditat de les persones necessitades.

El nou sistema tindrà capacitat per:

- Ésser utilitzat per tots els ciutadans.
- Entendre correctament la pregunta que se li formula.
- Identificar la pregunta i la resposta als repositoris de FAQs prèviament definits.
- Respondre al ciutadà amb la resposta encertada.

I tot això mentre el personal especialitzat prepara les respostes, i realitza tasques de manteniment i millora del sistema i concentra els millors esforços en la millora de la gestió i en una organització innovadora del servei.

2.4. Estudi d'alternatives per la solució

“La Caixa”

Aquí veiem com l'anomenat “assistent interactiu” de la Caixa, terme enganyós donat que no incorpora tecnologia agent ni intenta interpretar la pregunta. Simplement subministra les FAQ més similars, a més de creuar (continua i excessivament) informació comercial. Finalment hi ha un link per simplement informar quan la resposta no és correcta o no hi ha satisfacció per part de l'usuari. Segons fonts de e-laCaixa, l'empresa que va desenvolupar i manté aquest servei, el cost entre la llicència de software i el desenvolupament del projecte va superar els 600.000 euros. A pesar del seu elevat cost, el rendiment no és espectacular i entenem que no compleix amb les necessitats de l'atenció ciutadana en l'administració pública.

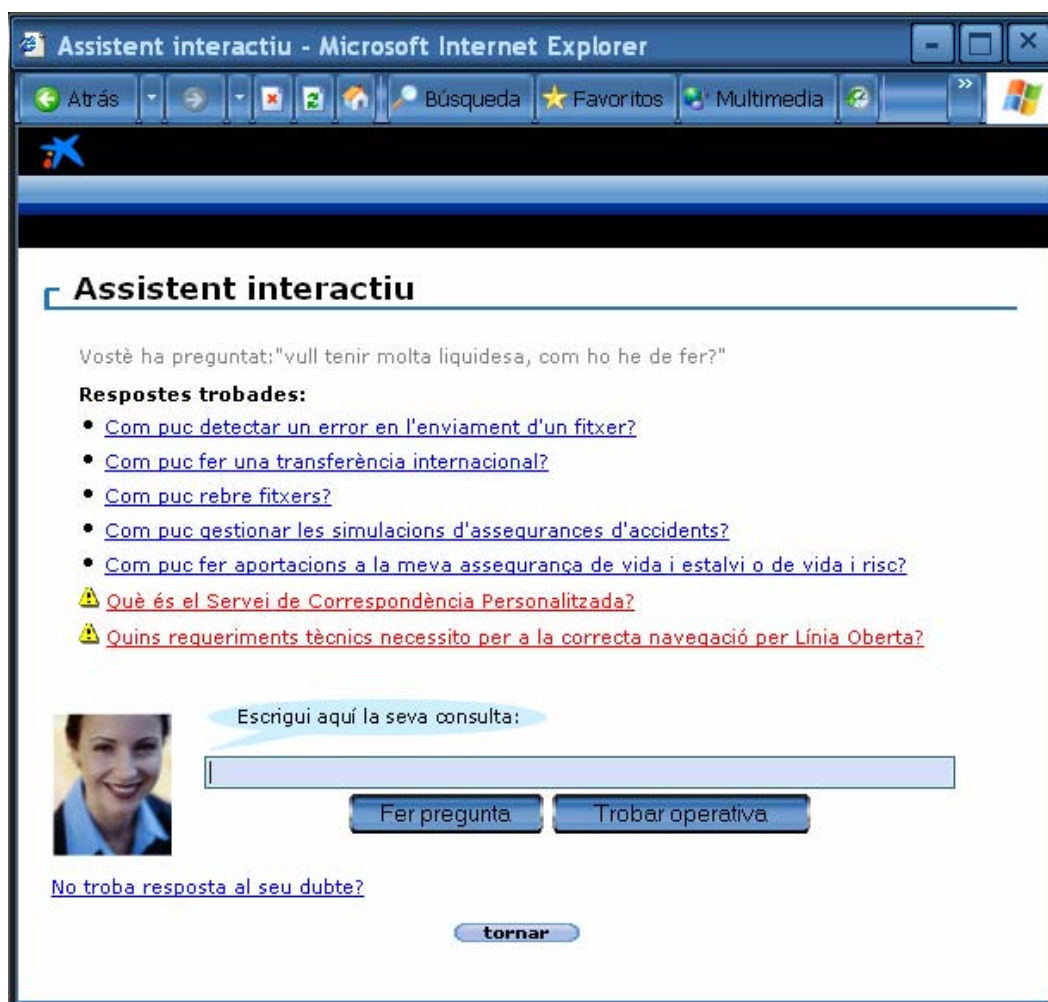


Figura 2: Alternativa de solució: “La Caixa”

Telefónica Espanya

En aquest altre exemple, veiem un altre assistent similar, amb una orientació encara més comercial, amb els mateixos defectes, exasperant davant tota pregunta que surti del guió comercial. Sabem que el seu cost va ser encara més elevat que el de “la Caixa”.

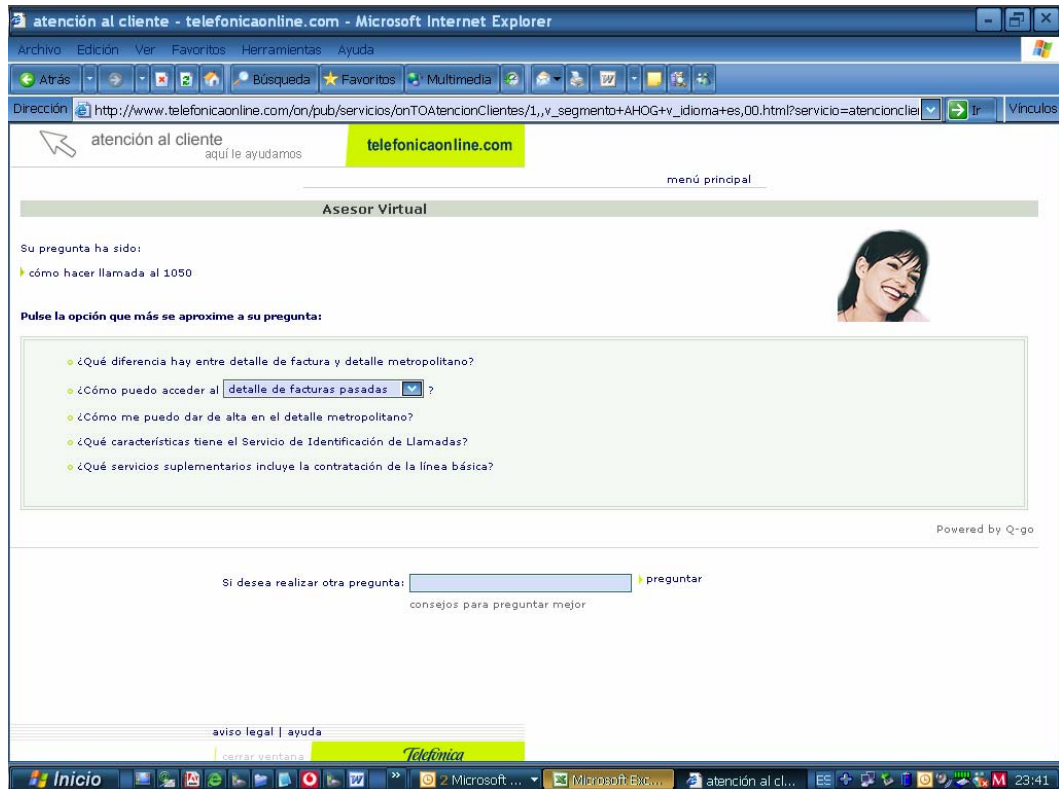


Figura 3: Alternativa de solució: Telefónica Espanya

QGo

De les eines del mercat de llenguatge natural, agafem QGo com exemple representatiu. QGo, el cost de la llicència de la qual oscil·la entre 120.000 i 200.000 euros segons la dimensió de l'empresa, requereix de projectes d'integració que van des de 100.000 fins a 400.000 euros addicionals, lo qual el converteix en un producte massa car per la immensa majoria d'administracions públiques.

El seu missatge és: “¿Cómo consigue usted el diálogo en línea significativo con sus clientes? Usando este software en línea para autoservicio, usted crea un ambiente de uso fácil y transparente donde los clientes y los usuarios internos pueden obtener igualmente respuestas a todas sus preguntas de una manera lógica. Este sofisticado software tiene una capacidad de aprendizaje para relacionar preguntas de los consumidores a la información que está siendo constantemente actualizada y refinada, basado en la entrada más última del mercado. QGo se asegura de que sus clientes reciben ayuda en línea en la forma de un servicio personal. Las respuestas dadas son por lo menos tan completas y de calidad como las que un agente del centro de llamadas podría ofrecer - a menos que sean mucho más fáciles, que entonces QGo responde más rápidamente y es 100 veces más barato manteniendo el nivel de las respuestas constantemente alto.”.

Tot i això, analitzant la seva capacitat de resposta veiem que és molt limitada, sense superar el baix nivell de les respostes aconseguides a “la Caixa” o “Telefónica”.

Sara

Mereix una menció especial, ja que es tracta d'un exemple al servei dels ciutadans en el seu contacte amb l'administració local de la ciutat de Malmö (Suècia) (<http://www.malmo.se>). El sistema ha estat desenvolupat per una firma sueca - Artificial Solutions- responsables també del robot Anna, de Ikea, entre d'altres. Es tracta d'un dels buscadors automàtics més potents i fiables entre els que existeixen en l'actualitat.

En paraules seves: “Gracias a nuestra competencia exclusiva en soluciones pedagógicas y manejables, combinada con los productos Kiwilogic, ofrecemos los Asistentes Interactivos más inteligentes del mercado.”.

Tot i això, les seves prestacions són molt limitades, no integrades en la base de la informació corporativa, per tant cares de mantenir i d'utilitzar (llicència d'un desenvolupament alemany), eficient en un catàleg reduït i tancat de productes o serveis, però molt difícilment adaptable a una administració pública, com és el cas de la ciutat sueca.

Les seves versions preliminars són altament insatisfactòries per una ciutat que ja disposi d'un bon sistema d'atenció ciutadana en funcionament al que els ciutadans no renuncien amb facilitat. I aquest és el de molts dels ajuntaments de Catalunya i de l'estat espanyol.



Figura 4: Alternativa de solució: Sara

Altres exemples representatius

En el camp dels assistents virtuals, existeixen nombrosos exemples, on podem destacar:

- Maria (<http://www.artificial-solutions.com>)
- Conversive (<http://www.conversive.com>)
- Anna de la firma Ikea
- (http://www.ikea.com/ms/es_ES/local_home/badalona.html)

Totes elles tenen en comú l'objectiu de subministrar conversa durant uns quants minuts per obtenir informació de l'usuari en llenguatge natural. Tot i això la seva implantació és molt limitada donat els seus frustrants resultats especialment per un ús recurrent en cerques d'informació.

2.5. Selecció de la solució

L'opció escollida finalment és la de realitzar una aplicació a mida, tenint en compte els aspectes més interessants de les altres alternatives estudiades i complint els requeriments.

Per diferenciar la contribució d'aquest projecte respecte a l'estat de l'art tant científic com tecnològic hem de ressaltar que el nostre sistema presentarà diferències significatives:

- Capaç de donar resposta inter administració. El sistema està especialment pensat per un entorn de l'administració pública catalana, espanyola, europea. Es defineixen entorns semàntics pel llenguatge administratiu modern i estàndard de l'administració pública.
- OpenSource i LGPL. A diferència dels productes existents actualment que són propietaris.
- Més econòmic, sense llicència, adaptable i escalable, adequat pel 99% dels ajuntaments espanyols. A diferència dels productes existents actualment, molt cars i només aptes per a grans ajuntaments com poden ser Barcelona, Madrid, València,...
- Més precís i més robust per donar la possibilitat d'entendre el context particular de cada ciutadà, mitjançant els agents aconselladors.
- Més eficient. Més fàcil de mantenir per part de l'administració. El sistema s'ajuda en les bases de dades existents en els sistemes d'informació de les corporacions públiques. Els ja existents creen una bossa de preguntes i una de respostes més o menys completes i previsibles, sempre molt limitades, que obliguen a mantenir la informació continguda de forma expressa i separada, per tant, redundant respecte al resta dels sistemes de organització.

Capítol 3. Anàlisi del sistema d'informació (ASI)

En aquest capítol s'exposa l'anàlisi del sistema d'informació, segons el que indica l'apartat de Mètrica v.3 que té el mateix nom. En aquest apartat, per un anàlisi orientat a objectes, la Mètrica v.3 proposa estudiar les següents activitats:

- Definició del sistema
- Anàlisi de classes
- Elaboració del model de dades
- Definició d'interfícies d'usuari

3.1. Definició del sistema

3.1.1. Determinar l'abast del sistema

Es vol automatitzar a través d'un portal web l'atenció al ciutadà per tal d'obtenir un servei més efectiu.

El nou servei web vol trencar les limitacions de personal i horari que hi han actualment, mentre que es millora també la informació proveïda i la personalització d'aquesta a les necessitats de cada ciutadà. Com a objectiu més ambiciós del projecte destaca la humanització del servei, és a dir, que el ciutadà se senti atès com si una persona estigués interactuant amb ell i responent a les seves qüestions.

Així doncs, aquest projecte pretén desplegar un software lliure de tecnologia de buscadors de FAQ (en anglès, Frequent Asked Questions) per oferir-lo a una ciutadania que sol·licita diàriament aquest servei. Aquestes FAQ es crearan des de l'experiència existent als serveis municipals d'Atenció Ciutadana.

En aquesta primera fase, el projecte es realitzarà especialment per les administracions de Catalunya i per tant en idioma català.

Els principals aspectes que s'han d'aconseguir en aquesta nova aplicació són:

- Oferirà una resposta preparada amb anterioritat per un expert, i per tant, la millor possible.
- Respondrà a preguntes inter administració que transcendeixin l'àmbit local i que per tant, són de difícil resposta a l'actualitat.
- Previndrà de demandes estacionals ocasionals. Oportunitat d'avançar informació.
- Facilitarà la alliberació de personal del front office per l'atenció automatitzada de les peticions habituals i repetitives.

- Facilitarà el dedicar més personal al back office (avaluació i gestió per assegurar el bon funcionament del front office, l'actualització del servei, la innovació i una major proactivitat).
- Facilitarà la derivació de més personal cap a les demandes complexes i a l'atenció de col·lectius amb necessitats especials.
- Reduirà el temps de formació del nou personal del servei, per l'experiència de la base de coneixement, i permetrà la seva incorporació més senzilla i ràpida a l'atenció personal.
- Ajudarà a definir millor els models d'atenció ciutadana, i crear estàndards que afavoreixin la seva evolució i sobretot la seva expansió als ajuntaments que encara no en disposin.
- Impulsarà la creació d'eines complementaries, estandarditzades i modulars, basades en software lliure i no propietària per ser implementades a qualsevol ajuntament o organització pública.
- Permetrà la creació de mòduls de formació col·lectiva i de reflexió sobre el model propi per ajudar als ajuntaments en la creació i la millora dels serveis d'atenció ciutadana.

Pel que fa a l'usuari:

- Entendre al ciutadà quan s'expressa en la seva llengua.
- Comprendre'l quan parla un llenguatge col·loquial, o no utilitza les paraules correctes per no conèixer l'argot de l'administració pública.
- Identificar la seva demanda encara que tingui errors d'ortografia, semàntica, idiomàtiques, gramaticals, tipogràfiques, ...
- Atendre al ciutadà en qualsevol demanda, encara que formuli preguntes que no transcendeixin a la responsabilitat de l'administració pública.

A continuació es mostren els principals processos que l'aplicació haurà de considerar:

- Gestió del sistema
- Cerca de FAQs

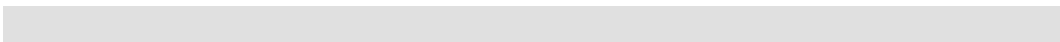


Diagrama de context del sistema

Un cop identificats els principals processos que l'aplicació ha de realitzar, es representa gràficament el sistema a partir del diagrama de context, identificant en aquest les entitats externes que hi formen part. Cada una d'aquestes entitats tenen els seus propis objectius i funcions a realitzar:

Administració: persones encarregades de proveir al sistema de la informació necessària perquè els usuaris puguin obtenir la resposta correcta a la seva pregunta.

Usuaris: persones encarregades de fer consultes i interactuar amb el sistema. També seran les encarregades de fer que el sistema aprengui.

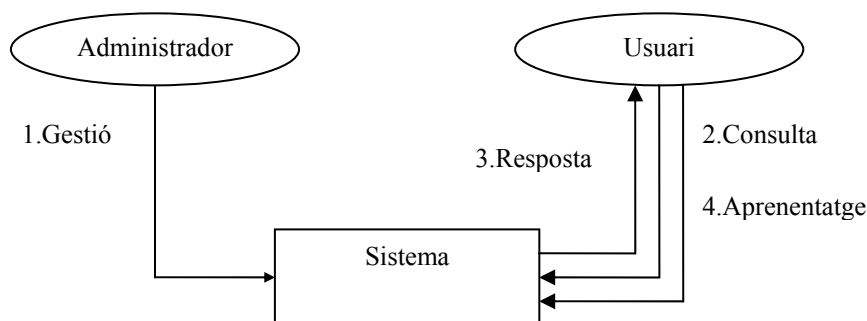


Figura 5: Diagrama de context del sistema

Els fluxos de dades entre el sistema i les entitats externes són els següents:

- **Gestió:** comporta la gestió de les FAQs, dels diccionaris, dels sinònims,... (no està inclòs dins l'abast del sistema)
- **Consulta:** conjunt de paraules que introdueix l'usuari per tal que el sistema li retorni la informació pertinent.
- **Resposta:** FAQ o llistat de FAQ's que retorna el sistema un cop introduïda la consulta per part de l'usuari.
- **Aprenentatge:** Interacció entre l'usuari i el sistema, que permet al sistema aprendre, i per tant millorar i afinar més en altres possibles cerques.

3.1.2. Procés de resposta en l'atenció automatitzada

El procés de resposta de iSAC en l'atenció automatitzada a la ciutadania és el següent:

- Comprensió correcta de la demanda que se li formula, en els termes propis de les persones del municipi, sense tecnicismes i en el propi idioma.
- Identificació de la pregunta a un repositori de FAQs dissenyat inicialment pels experts en l'Atenció Ciutadana.
- Validació davant del ciutadà que la pregunta s'ha comprès degudament, amb la formulació de la pregunta en els mateixos termes, senzills però més correctes, en que és recollida al repositori de FAQs.
- Identificació de la resposta més adequada al repositori de respostes a les FAQs, també prèviament definides pels experts del servei.
- Respondre al ciutadà amb la resposta encertada.
- Capacitat per expressar la satisfacció o no per la resposta rebuda i pel servei en general si és el cas.
- Capacitat per sobreposar-se a incorreccions ortogràfiques i morfològiques (barbarismes, castellanismes, estrangerismes,...).

La persona selecciona la pregunta que identifica millor la seva necessitat. A la mateixa pantalla l'hi apareix formulada correctament i completa la pregunta seleccionada, per a comprovació, i ja la resposta adequada, resposta completa preparada pel personal del servei, amb el temps necessari i havent pogut fer les consultes pertinents a personal especialitzat, si és el cas.

3.1.3. El motor de cerca iSAC

S'ha procurat crear un motor de cerca específic pel llenguatge de l'administració i d'aquesta manera, que atorgués un valor afegit a aquest sistema, que una altra eina de cerca no pogués fer al disposar d'un funcionament genèric.

Podem dividir el sistema en tres parts:

- La integració en els diversos sistemes d'informació comuns de l'administració pública
- La xarxa: base de coneixement de l'administració
- El motor de llenguatge natural



Integració dins l'administració pública

Tot i que no està a l'abast d'aquest projecte, iSAC haurà d'incorporar un mòdul d'integració, ja que aquest sistema no pot estar desconnectat del sistema d'informació existent de l'administració pública, per tal que de forma periòdica i automàtica, es generin noves FAQs, s'actualitzin les actuals o es completin amb informació complementària.

Els ajuntaments, consells comarcals,... ja posseeixen de bases de dades d'informació pels ciutadans. Per exemple, molts ajuntaments disposen de bases de dades amb informació sobre les activitats que s'organitzen a la ciutat, amb informació sobre tràmits o un directori de telèfons i adreces de les empreses i entitats de la ciutat. Com que aquestes bases de dades ja disposen d'un manteniment periòdic, és important que iSAC utilitzi aquesta informació actualitzada per tal de generar, actualitzar o completar les FAQs.

Per agafar com exemple l'Ajuntament de Terrassa, es tenen dades de diferents tipus, contexts:

- Directori
- Agenda
- Tràmits

A partir d'aquestes bases de dades, aquest mòdul, a partir d'ara distiller, haurà de transformar aquestes dades en les FAQs que retornarà el sistema quan l'usuari faci la consulta.

Es tracta d'un procés automàtic que a partir de les dades entrades, genera les FAQs. A més, el distiller crearà una taula de descriptors, on introduirà aquelles paraules o conjunt de paraules amb la seva importància dins la FAQ.

Posem un exemple pràctic. En cas de voler afegir una acte (agenda) al repositori de FAQs, s'afegirà a la nova taula de descriptors, el nom de l'acte amb un pes alt. El lloc on es realitza l'acte tindrà un pes mig. Pel contrari, l'organitzador de l'acte tindrà un pes baix.

Això passarà per tots els diferents camps de cadascun dels diferents contexts. Depenent de quin camp sigui, el seu valor tindrà un pes "Molt Alt", "Alt", "Mig" o "Baix" a la base de dades. Posteriorment aquests valors es quantificaran en valors reals.

Xarxa de coneixement de l'administració pública

És molt important que el motor de llenguatge natural incorpori un coneixement lèxic i semàntic propi de cada administració pública. Per tal fer més efectiu el procés de recerca i humanitzar la interacció amb l'usuari és necessari que el sistema incorpori un coneixement propi de l'ajuntament del municipi, administracions públiques en general, noms propis de la ciutat,... En resum, coneixement de base que hauria de tenir qualsevol empleat del servei d'atenció ciutadana.

Tot aquest coneixement s'utilitza juntament amb el motor per tal d'identificar el quina és la resposta a la pregunta realitzada i també per tal de personalitzar la resposta a l'usuari i donar-li més informació que li pugui ser útil.

Tota aquesta informació s'afegirà a una taula de sinònims, per tal de tenir en compte al realitzar la cerca.

Algorisme de cerca

La següent figura ens mostra la interfície que veu un ciutadà quan accedeix al servei i des del qual s'iniciarà tot el procés de tractament, avaluació, cerca i presentació de resultats, a partir de la cadena de text entrada.



Figura 6: Pantalla d'inici

La interfície és molt senzilla, no es vol donar peu a possibles confusions, per això es limita a un quadre de text en el qual el ciutadà introduirà la seva petició, i un botó que s'utilitzarà per enviar la cerca al sistema i començar el procés.

PAS 1 – Entrar la cadena de cerca: El pas imprescindible, evidentment, és la cerca que fa el ciutadà i que envia al sistema. Aquesta captura és senzilla, es limita a rebre la cadena de lletres o el conjunt de paraules a través de la xarxa de internet sense fer-hi cap modificació. D'aquesta manera no alterem la forma ni l'estructura i, tot i que acte seguit la tractarem, es conserva per poder fer una nova cerca amb les mateixes paraules que el ciutadà ha escollit.

Es seguirà la política de no modificació de la cadena de cerca. Aquesta política agafa consistència si tenim en compte el cas que s'introdueixen paraules que nosaltres podríem arribar a considerar incorrectes, però que en el context d'alguna FAQ són perfectament vàlides. Podria ser el cas del nom d'alguna entitat social, per exemple; és possible que no s'hagi tingut en compte aquest nom al tractar la cerca, però no per això la considerem incorrecta i la descartem com a tal.

Aquesta cerca la podem realitzar a totes les FAQs en general, o reduir la cerca cap a un context determinat, reduint així el temps de resposta i el nombre de possibles solucions per part del sistema. La forma de reduir aquesta cerca és escollint el context que té la nostra consulta mitjançant la barra d'eines iSAC.

PAS 2 – Pretractament inicial: Un primer pas a realitzar amb aquesta cadena entrada, independent de la resta, és treure els accents introduïts, així com transformar totes les paraules a lletra minúscula. També s'eliminaran els números, excepte aquells que s'interpreta que poden aportar informació, com podria ser un número seguit d'un mes de l'any, interpretant el sistema que es tracta d'una data.

PAS 3 – Correccions ortogràfiques: Un procés en paral·lel que es realitza amb aquesta cadena entrada, independent de la resta, és la proposta de correccions ortogràfiques a partir de la cerca inicial realitzada per l'usuari del sistema. Aquesta correcció ortogràfica prova d'analitzar tot el text contingut a la cadena de cerca i trobar aquelles paraules que no són acceptades a la llengua catalana, es poden donar, però, diversos casos:

- Un cas seria el d'una paraula mal escrita però que està contemplada dins de la nostra base de dades com a “paraules o noms propis” relacionats amb alguna o algunes FAQs. Per tant aquesta ha de ser considerada com a correcte.
- L'altre cas és el cas més comú dins el que serien les correccions ortogràfiques, una paraula mal escrita, el tractament que en farem es lleugerament diferent al que es fa normalment a les correccions. El sistema tindrà en compte dos barems a l'hora de triar la millor correcció possible d'una paraula, el primer és el de la proximitat o semblança morfològica de la paraula “mal escrita” i de la proposta de correcció, el segon és el major grau d'ús de la paraula proposada dins el món de l'administració, de manera que la paraula que obtingui un valor més gran al combinar aquests dos classificadors serà la triada per a la proposta de correcció ortogràfica.

Un tema a tenir en compte respecte a la correcció ortogràfica, és que no es valorarà com a paraula mal escrita tota aquella que hagi de portar signes ortogràfics i no en porti o els porti de forma incorrecta, aquests s'eliminen donat que al sistema no li importen al no aportar significat extra.

La correcció ortogràfica, com ja s'ha dit, és independent de la resta del procés, tant del tractament que es farà de la cerca, com de la pròpia cerca de FAQs coincidents. Aquest procés es limita a fer una proposta de correcció ortogràfica i l'usuari del sistema triarà si l'aplica o si opta per no modificar el text original.

Per tal de millorar en eficiència amb el tema de les correccions, i poder estar a l'últim dia pel que fa a noves paraules incorporades als diccionaris de cada llengua, s'han utilitzat els diccionaris de lliure distribució de l'aplicació OpenOffice. A més aquests diccionaris ens ajudaran també alhora de la lemmatització, ja que ens indica per cada paraula si es tracta d'un adjectiu, conjunció, determinant, interjecció, nom, pronom, adverb, preposició verb, nom propi o abreviatura.

El sistema, en el cas d'haver-hi un error ortogràfic, proposarà sempre la solució més popular, és a dir, aquella paraula que té més opcions segons el diccionari de ser la paraula correcta.

A la figura es veu com el ciutadà ha introduït la paraula "cinmes", i el sistema li proposa una correcció de tota la cadena de cerca sencera. Tot i així s'ha realitzat la cerca amb les paraules introduïdes i es mostren els resultats obtinguts amb aquestes paraules.



Figura 7: Exemple de correcció ortogràfica

PAS 4 – Eliminació de stopwords: El tractament pròpiament de la cadena de cerca introduïda pel ciutadà comença amb l'eliminació de les “stopwords” o paraules freqüents. Aquestes paraules freqüents representen tot aquells pronoms, determinants, articles i qualsevol altra paraula que, donat el seu ús massa freqüent o de la seva funció sintàctica, no aporten cap valor afegit al significat global de la cerca. Aquesta discriminació és molt senzilla, es disposa d'un llistat de paraules a bloquejar, que pot editar el responsable del sistema, i es comprova que no hi hagi cap paraula de les existents a aquesta llista dins la cerca feta pel ciutadà. En el cas que n'hi hagi alguna, aquesta s'eliminarà i es procedirà a continuar amb al següent pas del tractament del text.

PAS 5 – Identificar el lexema: Un cop s'han eliminat totes aquelles paraules freqüents de la cerca, es procedirà a fer la lemmatització, o extracció del lexema, de totes les paraules que han superat el primer filtre.

L'objectiu d'aquest pas és el d'eliminar aquella informació que no ens és útil i simplement conservar la que ens portarà a una millor obtenció de resultats.

Així d'una paraula com “gossos” o “gosses” només ens interessa el seu lexema: “gos”. La resta no ens aporta una informació el suficientment útil com per conservar-la. Recordem que igualment es conserva la cerca inicial sense cap modificació per personalitzar i realitzar millor l'acotament dels resultats.

En el cas dels verbs, aquest procés ens extrauria el temps verbal, la persona, el gènere,...

Això ens és molt útil per tal de no haver d'introduir dins els descriptors de les FAQs totes les possibles conjugacions i variants de cada paraula, fet que seria molt costós, tant a nivell d'esforç per part del servei i l'administrador, com per volum de dades al nostre sistema.

Sense aquest procés de lemmatització, aquesta resposta no hauria sigut vàlida pel sistema.

Per tal de millorar en eficiència amb el tema del lexema, s'ha utilitzat el diccionari de lliure distribució del FreeLing.

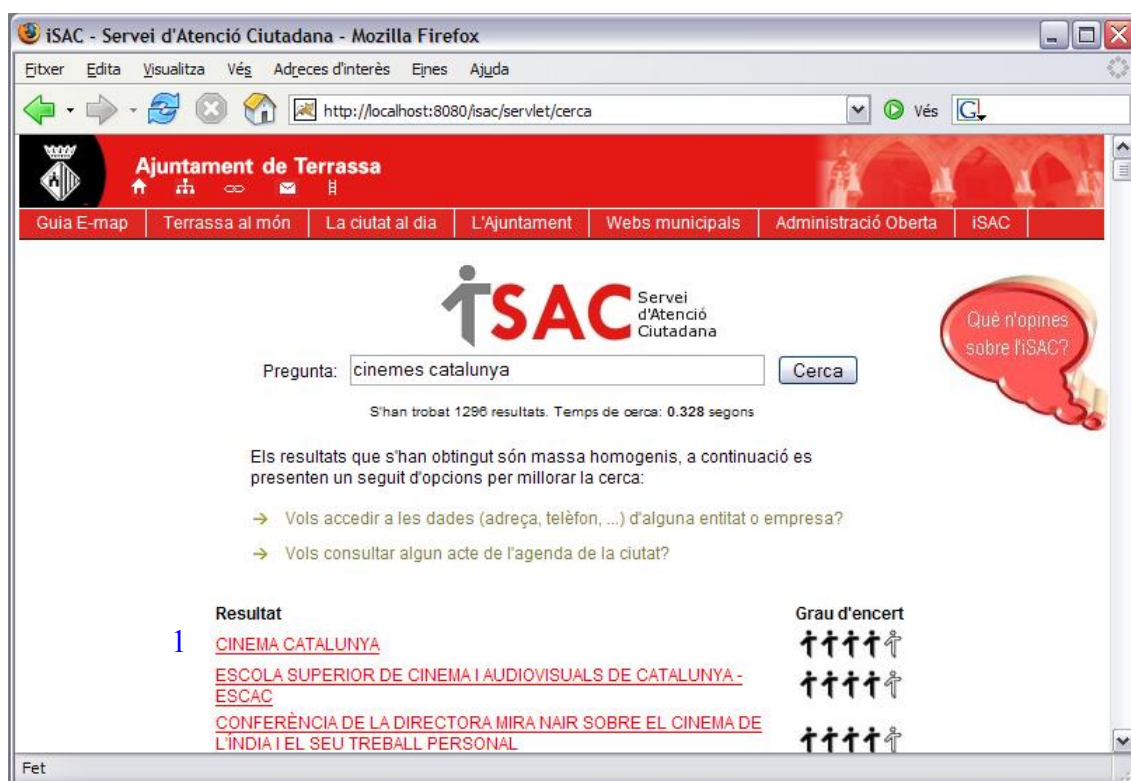


Figura 8: Resultats del sistema havent lemmatitzat

A la figura es pot veure com havent introduït la paraula “cinemes catalunya”, una de les possibles respostes retornada pel sistema ha estat (1) “cinema catalunya”, on no hi figura aquesta paraula.

PAS 6 – Cerca de sinònims: La cerca de sinònims és un dels passos interns que realitza el sistema, sense que l'usuari hi pugui intervenir. A partir de la cadena de cerca, el sistema busca sinònims de cadascuna de les paraules entrades per tal d'ampliar i millorar la cerca.

Aclarir que la cerca de sinònims dins la taula esmentada es realitza mitjançant la comanda SQL “LIKE”. Amb això ens assegurem també que possibles expressions escrites mínimament diferent també apareguin com a sinònim.

Aquesta cerca de sinònims és molt més que això. El sistema examina dins la taula de sinònims si alguna de les paraules introduïdes pel sistema pot tenir alguna altra paraula associada dins la taula de sinònims que tenim a la base de dades.

Dins aquesta taula hi tindrem, a més dels sinònims pròpiament dits, barbarismes, expressions populars, expressions típiques del llenguatge verbal com pot ser parlar de d'un dia sense especificar la data (Nadal, 25 de desembre), informacions locals (prominent de la xarxa de coneixement de l'administració ciutadana), dialectes de la zona,...

Quan s'introdueixin “expressions de temps”, el que el sistema farà serà transformar-les en dates. Per exemple, si introduïssim la consulta “qui actual al concert de demà”, el sistema buscaria el concerts (o altres actes) que es produïssin en la data en la qual correspongués a “demà”.

El fet d'analitzar sinònims beneficia la llibertat d'expressió i la manera d'entrar cada ciutadà la seva consulta. Dos usuaris poden voler el mateix i poden realitzar la consulta utilitzant paraules diferents, però el sistema respondre el mateix per ambdós casos.

PAS 7 – Identificar el context: Un filtre que intenta aplicar el sistema, per intentar acotar el rang, és intentar identificar el context de la cerca (en el supòsit que l'usuari no l'hagi introduït anteriorment des de la barra d'eines iSAC). En cas que el context de la cerca es redueixi a un, implicarà que el sistema únicament buscarà dins aquell context de FAQ. Si no és així, recordarà el percentatge de correspondència amb cada context per tal d'utilitzar-lo més endavant, i realitzarà la cerca per tot el conjunt de FAQs.

PAS 8 – La cerca: Finalment, ja només ens queda fer la cerca de cadascuna de les paraules (ja siguin les escrites per l'usuari com les afegides pel sistema) dins la taula de descriptors.

Segons el nombre de paraules coincidents a cada FAQ, el pes que tenen aquestes dins d'ella, així com els percentatges de context de les paraules analitzat anteriorment, ens serviran per retornar un llistat de FAQs ordenat per afinitat a la consulta entrada.

El càlcul d'aquesta afinitat es calcula mitjançant la següent fórmula:

$$[1 - (0,35^{\text{SUM}(\text{Pes descriptors coincidents})}) * (\text{número de descriptors coincidents} / \text{número de paraules identificades})] * (90\%) + (10\% \text{ Pes de la FAQ})$$

Posteriorment, en cas que la cerca es realitzi per més d'un context, aquest valor serà el 90%, essent el 10% restant el percentatge de possibilitat del context calculat en el pas anterior.

El fet d'utilitzar aquest valor 0,35, es deu al fet que es buscava obtenir una gràfica asimptòtica que tendeixi cap al valor 1. A més, el pendent aconseguit és molt satisfactori.

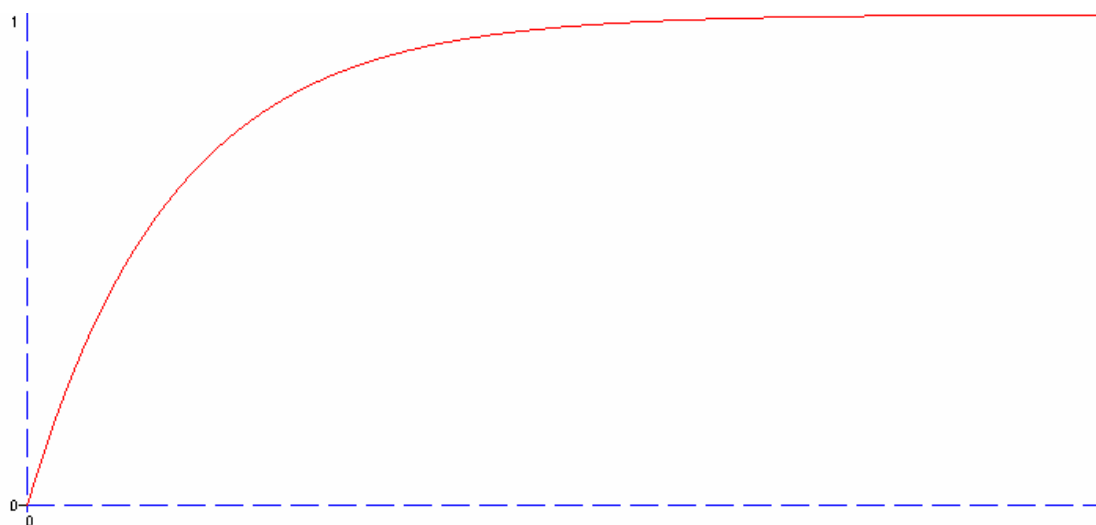


Figura 9: Gràfica corresponent a la fórmula del càlcul d'afinitat

Amb aquest sistema es valora més la quantitat de coincidències que no pas la qualitat. Interpretem que és millor que coincideixin moltes paraules de la consulta (tant la inicial com amb les transformacions que ha generat el sistema) amb els descriptors de la FAQ, que no que aquestes siguin exactament les introduïdes únicament per part de l'usuari.

En cas d'empat entre dues o més FAQs a percentatge de possible resultat, el sistema els mostrarà aleatòriament per tal de no afavorir sempre una possible mateixa resposta.

PAS 9 – El refinament: Hi ha diversos casos que ens podem trobar alhora de realitzar una cerca:

- No hi ha cap resposta o FAQ que tingui coincidències amb la cerca proposada.
- Hi ha una o diverses respostes o FAQs i una és prou bona com per ser mostrada de forma automàtica.
- Hi ha una o diverses respostes, no hi ha cap de bona però hi han coincidències amb els “refinaments” controlats.
- Hi ha una o diverses respostes, no hi ha cap de bona, no hi han coincidències i el rànquing general és prou elevat.
- Hi ha una o diverses respostes, no hi ha cap de bona, no hi han coincidències i el rànquing general és baix.

Segons cada un dels casos s'actuarà d'una forma diferent, aplicant a cada cas el mètode òptim.

La forma com el sistema s'enfrontaria als casos trobats, és la següent:

- Pel primer cas, quan no hi ha cap resposta o FAQ que tingui coincidències amb la cerca proposada, el sistema emetrà un missatge del tipus: “Ho sento però no hi han respostes al sistema per la qüestió proposada.”, mostrant, de la mateixa manera, un enllaç que permeti al ciutadà posar-se amb contacte via correu electrònic amb el servei convencional d'atenció ciutadana.
- En el cas que hi hagi un resultat prou bo per ser mostrat de forma automàtica, independentment del nombre de resultats total obtingut, llançarà la pàgina web per tal que el ciutadà pugui veure la informació requerida, permetent en tot moment la opció de feedback i valoració del resultat obtingut.

La manera de saber si aquest resultat és prou bo com per respondre directament, és observant si la possible resposta té un pes més gran de 0,5 (50%), i a més, la diferència amb la següent solució amb més pes de la llista de resultats, és major a 0,3 (30%). En aquest cas el sistema retornarà la primera solució únicament de forma automàtica.

- Si el que ha passat és que hi ha un o diversos resultats, cap de bo, però hi ha coincidències amb el tipus de “refinament” tipificat, s’aplicarà el refinament oportú:
 - Preguntes del tipus directori (activitats, entitats tipus banc, botigues,...) és proposaran diverses opcions de “refinament”:
 - Escollir un barri/carrer concret.
 - Mostrar un llistat de les entitats que desenvolupin l’activitat.
 - Preguntes del tipus agenda es proposarà:
 - Escollir l’organitzador de l’acte.
 - Poder triar la categoria que es desitja.

En qualsevol dels casos és mostrarà el llistat de resultats obtinguts amb la cerca proposada.

- Si hi ha una o diverses respostes o FAQs, no hi ha cap de bona ni cap refinament relacionat, però el rànking té uns valors acceptables, no es proposarà cap mena de refinament, i simplement és mostrarà el llistat de resultats obtinguts.
- Per últim, si hi ha diverses preguntes o FAQs, cap de bona ni cap refinament relacionat, i el rànking general és baix, és proposarà un tipus de “refinament” bàsic, és a dir, es proposaran les següent preguntes o algunes de semblant que encaminaran la pregunta cap a un context concret:
 - Vols obtenir la direcció d’una entitat o negoci?
 - Estàs interessat en informar-te sobre alguna activitat de l’agenda ciutadana de la ciutat?
 - Proves d’accedir o rebre informació sobre algun tràmit legal?

A banda del “refinament”, en aquest cas també és mostrarà sempre el llistat de resultats obtingut.

A la següent figura podem veure un esquema d’aquest refinament.

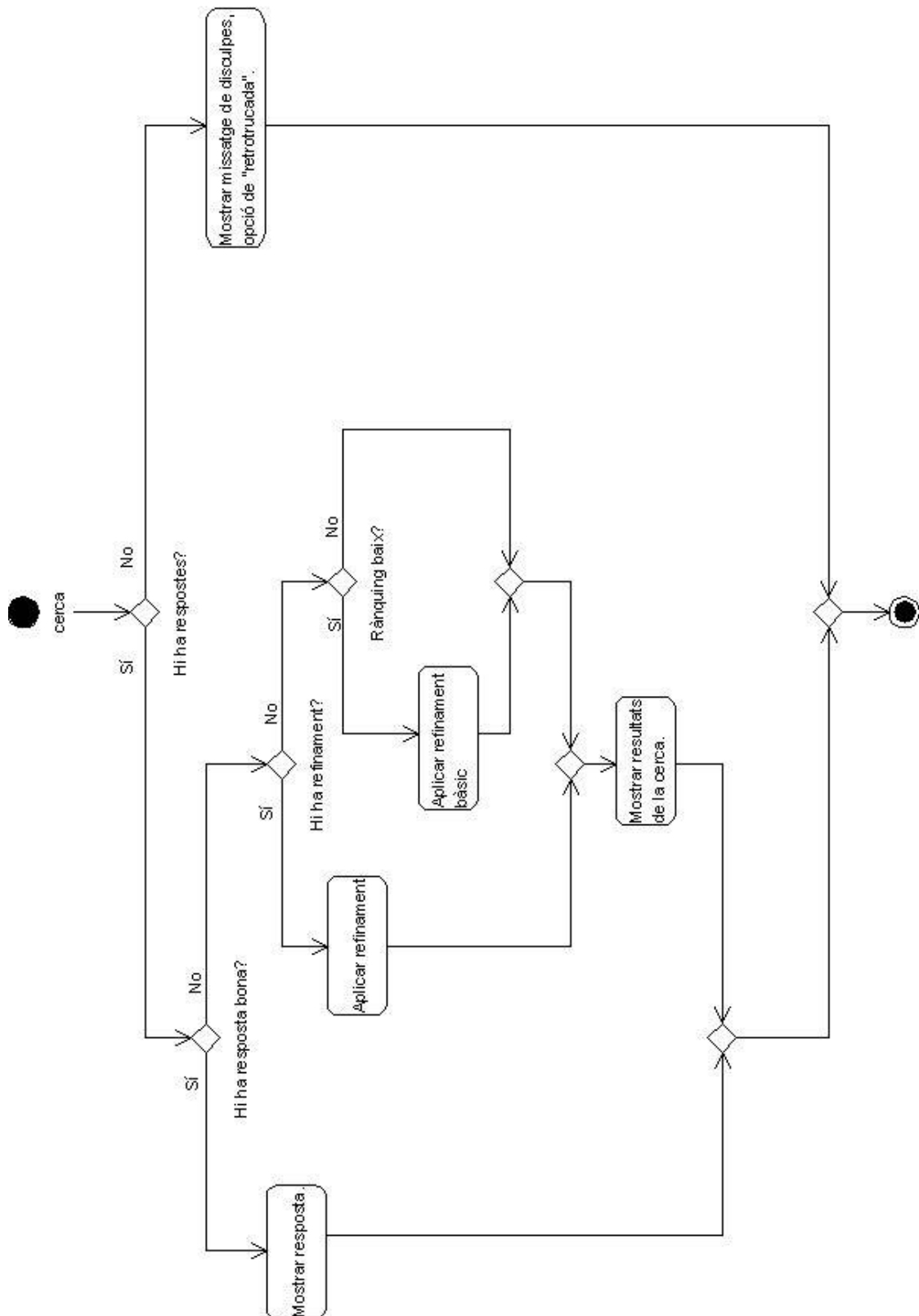


Figura 10: Esquema del refinament

PAS 10 – Aprenentatge: Un cop acabada la cerca, el sistema aprèn. Aquesta manera d'aprenentatge és de dos tipus:

- Aprenentatge per la correcció: com ja s'ha comentat, en el cas d'haver-hi un error ortogràfic, el sistema proposarà com a correcció ortogràfica la paraula més utilitzada (sempre dins el llistat de possibles solucions que retorna el diccionari). Per tant, aquest aprenentatge, ens servirà per actualitzar les paraules utilitzades i per tant millorar les possibles correccions posteriors.
- Aprenentatge de les FAQs: El fet d'escollir una FAQ, farà que el sistema afegeixi valor a aquella FAQ per a properes consultes, ja que el seu "índex de popularitat" ha incrementat.

3.2. Establiment de requisits

3.2.1. Obtenció de requisits

Prenent com a punt de partida els requeriments obtinguts en l'apartat de la definició dels requisits i dels casos d'ús d'alt nivell de l'apartat de la determinació de l'abast del sistema, a continuació es descriuen amb més detall els requeriments segons els següents tipus:

- Funcionals
- Rendiment
- Seguretat
- Disponibilitat del sistema

Requeriments funcionals

Per tal d'analitzar amb més detall els requeriments funcionals, s'estudiaran els casos d'ús associats a aquests:

- Barra eines Internet Explorer:
 - Actor principal: usuari
 - Descripció: Podrà instal·lar les barres d'eines pel navegador Internet Explorer.
- Barra eines Mozilla Firefox:
 - Actor principal: usuari
 - Descripció: Podrà instal·lar les barres d'eines pel navegador Mozilla Firefox.

- Cercar:
 - Actor principal: usuari
 - Descripció: L'usuari podrà introduir la seva consulta al sistema.

- Acceptar Correcció:
 - Actor principal: usuari
 - Descripció: Es pot donar el cas que l'usuari hagués introduït una paraula incorrecta. El sistema li proposarà una correcció i ell haurà de decidir si accepta la correcció o no.

- Llistar Resultats:
 - Actor principal: usuari
 - Descripció: El sistema retornarà un llistat amb tots els possibles resultats, ordenats per ordre d'afinitat entre la consulta realitzada i el llistat de FAQs del sistema. En cas de no trobar cap resultat, mostrarà un llistat buit, o un missatge indicant de la incidència.

- Escollir Resultat:
 - Actor principal: usuari
 - Descripció: Un cop mostrat el llistat amb totes les FAQs que el sistema troba semblances, l'usuari haurà d'escollir aquella que considera que s'aproxima més al que vol.

- Mostrar Resultat:
 - Actor principal: usuari
 - Descripció: És la FAQ escollida per l'usuari com aquella que considera millor pel que vol consultar. Se li mostrarà la informació de la FAQ. També es pot donar el cas que la consulta introduïda per l'usuari, solament tingui una possible resposta per part del sistema, i per tant ja la mostrarà directament, sense passar per aquest pas anterior d'haver de llistar resultats.

Per tal de comprendre millor l'especificació dels casos d'ús exposats, a continuació es mostra el diagrama de casos d'ús associat:

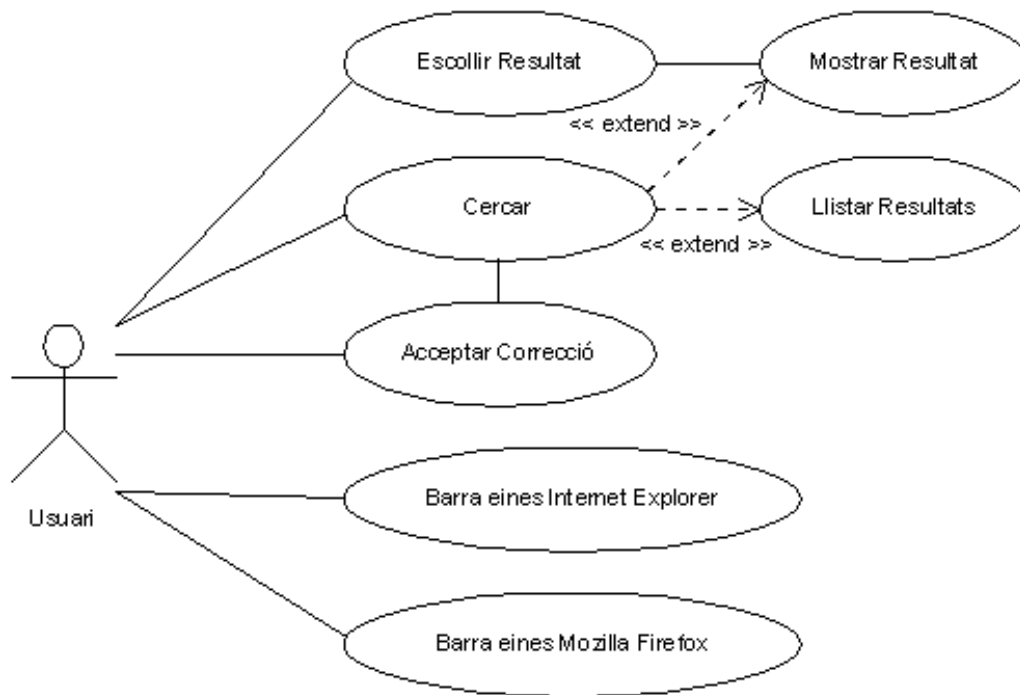


Figura 11: Diagrama de casos d'ús

Requeriments de rendiment

El rendiment de l'aplicació a construir ha de ser suficient per tal de poder dur a terme els requeriments funcionals demanats. Tant sols es demana que pugui ser executat amb garanties en un ordinador estàndard del mercat actual.

Requeriments de seguretat

El sistema ha de garantir la consistència de les dades i la seguretat d'aquestes. Per tant caldrà que es disposi d'un sistema de realització de còpies de seguretat.

Requeriments de disponibilitat del sistema

El sistema ha d'estar disponible per qualsevol persona que el vulgui utilitzar. Per aquest motiu no cal cap gestió d'usuaris ni control d'accessos a l'aplicació.

Únicament existirà l'usuari administrador, que serà l'encarregat de poder gestionar (a través de la web) els diferents diccionaris, tot i que aquest aspecte no entra a l'abast del projecte.

3.2.2. Especificació de casos d'ús

Aquest apartat té com a objectiu fer una especificació dels casos d'ús trobats en el punt anterior. Aquesta especificació fa referència a:

- Descripció de l'escenari, com l'usuari interactua amb el sistema i quina és la resposta que n'obté.
- Precondicions i postcondicions.
-
- Barra eines Internet Explorer:
 - Interacció: L'usuari podrà instal·lar la barra d'eines del navegador clicant sobre la icona on hi ha la imatge del seu navegador.
 - Precondicions: No hi han precondicions.
 - Postcondicions: S'ha clicat sobre la icona del navegador utilitzat .
- Barra eines Mozilla Firefox:
 - Interacció: L'usuari podrà instal·lar la barra d'eines del navegador clicant sobre la icona on hi ha la imatge del seu navegador.
 - Precondicions: No hi han precondicions.
 - Postcondicions: S'ha clicat sobre la icona del navegador utilitzat .
- Cercar:
 - Interacció: L'usuari introduirà la seva consulta a la caixa de text. Un cop escrita clicarà sobre el botó "Cerca" per tal de procedir amb la cerca de resultats.
 - Precondicions: No hi han precondicions.
 - Postcondicions: La única postcondició és que la cerca no pot ser buida. En cas de ser-ho, el sistema no farà res.

- **Acceptar Correcció:**
 - **Interacció:** Si el sistema retorna que ha trobat una incorrecció ortogràfica, l'usuari podrà senyalar si vol acceptar o no la proposta ortogràfica que el sistema li suggereix. Únicament haurà de clicar sobre un hipervincle si decideix acceptar-la.
 - **Precondicions:** L'usuari ha introduït una paraula que no es troba al diccionari de paraules.
 - **Postcondicions:** No hi ha postcondicions.
- **Llistar Resultats:**
 - **Interacció:** El sistema mostrarà tots els resultats trobats. A més l'usuari podrà tornar a introduir una altra cerca en cas que ho consideri convenient.
 - **Precondicions:** S'ha trobat més d'una solució o FAQ que satisfà la cerca.
 - **Postcondicions:** No hi ha postcondicions.
- **Escollir Resultat:**
 - **Interacció:** L'usuari clicarà sobre una de les respostes proposada pel sistema. Haurà d'escollir quina d'elles s'assembla més al motiu de la seva consulta.
 - **Precondicions:** El sistema ha mostrat més d'una solució.
 - **Postcondicions:** No hi ha postcondicions.
- **Mostrar Resultat:**
 - **Interacció:** En aquest cas l'usuari podrà interactuar de dues maneres amb el sistema. Una d'elles seria llegint, anotant, memoritzant,... la informació mostrada. Una altra manera seria clicant sobre algun possible link que aparegui als resultats. L'usuari també podrà tornar enrera i tornar a visualitzar la llista de possibles resultats.
 - **Precondicions:** El sistema ha trobat una possible resposta o bé l'usuari ha clicat sobre l'opció desitjada.
 - **Postcondicions:** No hi ha postcondicions.

3.3. Anàlisi dels casos d'ús

Durant l'anàlisi dels casos d'ús cal identificar les classes associades a cada un d'ells i la interacció entre els diferents objectes. És en aquest apartat on s'exposen aquests punts.

3.3.1. Identificació de classes associades a un cas d'ús

Tenint en compte els models i especificacions dels casos d'ús identificats en l'apartat de l'establiment de requisits, a continuació es mostren les classes que formen part de cada cas d'ús:

- Barra eines Internet Explorer: en aquest cas d'ús no intervé cap classe.
- Barra eines Mozilla Firefox: en aquest cas d'ús no intervé cap classe.
- Cercar: intervenen totes les classes, ja que ha de buscar que les paraules estiguin dins el diccionari (PARAULA), extreure paraules freqüents (STOPWORD) i després completar la cerca a partir dels sinònims (SINONIM) i els lexemes (LEXEMA i LEXEMATIPUS). Un cop fet anirà a buscar coincidències amb les FAQs (GRUP, DESCRIPTORS i RESPOSTA).
- Acceptar Correcció: no intervé cap classe, ja que la possible correcció ja l'ha mostrat i si l'usuari decideix acceptar-la, es tornarà a realitzar una altra cerca.
- Llistar Resultats: Intervé la classe RESPOSTA, ja que mostrarà el títol de totes aquelles FAQs que ha trobat coincidències.
- Escollir Resultat: El fet d'escollir un resultat no intervé cap classe.
- Mostrar Resultat: Intervé tant la classe RESPOSTA, ja que es mostren les dades de l'opció escollida per l'usuari o bé de l'única opció trobada pel sistema.

3.4. Anàlisi de classes

LEXEMATIPUS: conté tots els tipus de paraules que existeixen i els hi assigna un identificador. Aquests tipus poden ser: abreviatura, adjectiu, adverb, conjunció, determinant, interjecció, nom, nom propi, preposició, pronom o verb.

LEXEMA: diccionari amb totes els lexemes. Extret del diccionari Freeling. A més s'identifica de cada paraula, de quin tipus es tracta segons la classe LEXEMATIPUS..

PARAULA: inicialment tindrà totes les paraules del diccionari OpenOffice de la corresponent llengua.

LEXEMAPARAULA: és la classe resultant de la relació entre la classe LEXEMA i la classe PARAULA. Tindrà identificat per totes les paraules el seu lexema.

SINONIM: contindrà tot un seguit de paraules que seran equivalències d'altres de la classe PARAULA.

STOPWORD: hi haurà tot un seguit de paraules que el sistema extraurà i no contemplarà alhora de realitzar la cerca. Aquestes bàsicament són determinants, articles,...

RESPOSTA: serà la classe encarregada de tenir totes les FAQs. En ella és d'on s'extraurà el resultat de la cerca per tal de mostrar-lo a l'usuari. A més tindrà un pes associat a cada FAQ.

DESCRIPTORS: serà aquell conjunt d'expressions i paraules que identifiquen cada FAQ. Serà on es realitzarà la cerca introduïda per l'usuari.

GRUP: conté tots aquells tipus de contexts que hi poden haver. Cada FAQ tindrà associat un d'aquests contexts.

A continuació podem veure com seria el diagrama d'aquestes classes:

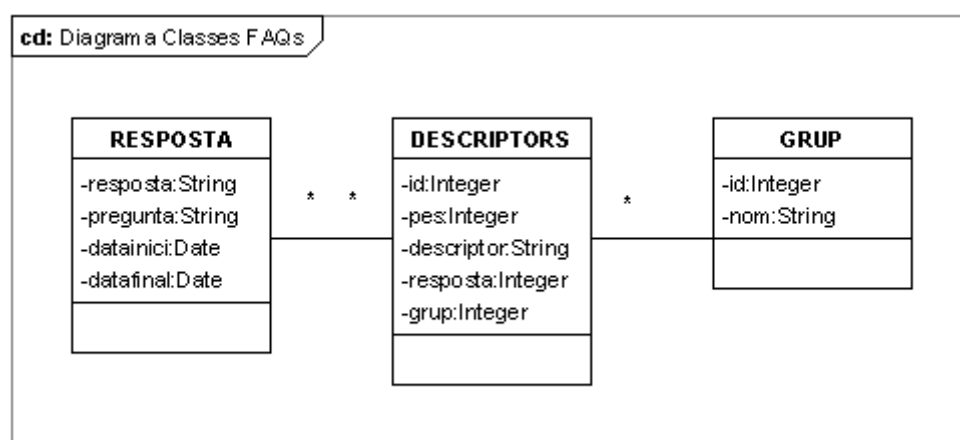


Figura 12: Diagrama de classes de les FAQs

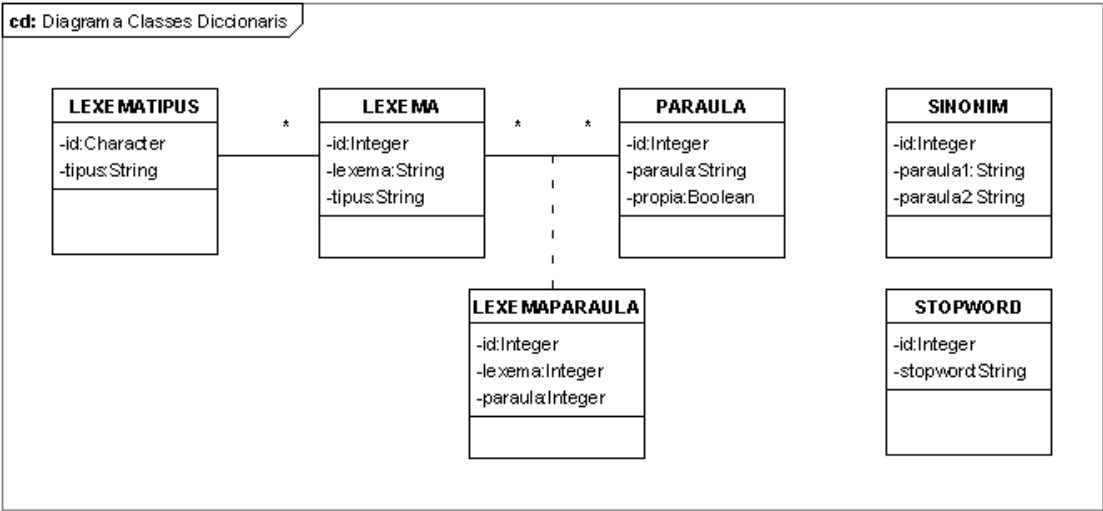


Figura 13: Diagrama de classes dels diccionaris

3.5. Elaboració del model de dades

3.5.1. Elaboració del model lògic de dades

Una vegada vistos els requisits que ha de complir l'aplicació i identificats els subsistemes s'ha d'elaborar el model de dades, identificant les entitats que queden dins de l'àmbit del sistema d'informació, els atributs de cada entitat, diferenciant aquells que poden convertir-se en identificadors de l'entitat.

En primer lloc es fa el diagrama Entitat/Relació, que es pot veure en els següents diagrames, i a continuació es fa el disseny lògic, en el qual ja es podran veure les diferents taules i camps per a la base de dades de l'aplicació.

Diagrama Entitat / Relació

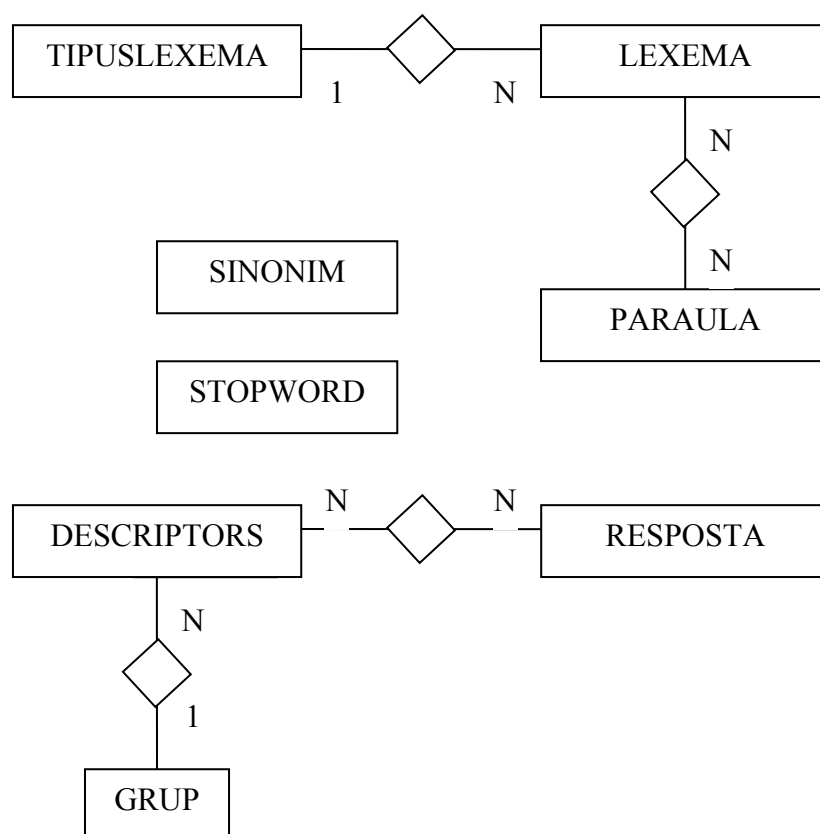


Figura 14: Diagrama entitat / relació

Model relacional

A continuació es poden veure els diferents atributs que componen cada una de les taules obtingudes a partir del model Entitat/Relació.

LEXEMATIPUS:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
tipus	tipus de lexema

LEXEMA:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
lexema	lexema
tipus	clau forana de la taula LEXEMATIPUS identifica el tipus de lexema que es tracta

PARAULA:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
paraula	paraula
propia	booleà que indica si es tracta d'una paraula pròpia del diccionari o s'ha afegit per part de l'administrador.

PARAULALEXEMA:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
lexema	clau forana de la taula LEXEMA identifica el lexema que té
paraula	clau forana de la taula PARAULA identifica la paraula

SINONIM:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
paraula1	paraula
paraula2	sinònim de la paraula1

STOPWORD:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
stopword	paraula (determinant,...)

RESPOSTA:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
resposta	explicació de la FAQ
pregunta	títol original de la FAQ
datainici	data que s'inicia la FAQ
datafinal	si es dona el cas, data en que la FAQ deixa de tenir sentit
pes	importància de la FAQ

DESCRIPTORS:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
pes	importància de la pregunta
descriptor	paraula o conjunt de paraules descriptors de la FAQ
resposta	clau forana de la taula RESPOSTA identifica la FAQ a la qual pertany
grup	clau forana de la taula GRUP identifica el context al qual pertany

GRUP:

<i>Atribut</i>	<i>Descripció</i>
id	identificador
grup	context (Agenda, Tràmit, Directori, Altres)

3.6. Definició d'interfícies d'usuari**3.6.1. Especificació de principis generals de la interfície**

La idea bàsica de la interfície és que ha de ser molt senzilla, no es vol donar peu a possibles confusions, per això es limita a un quadre de text en el qual el ciutadà introduirà la seva petició, i un botó que s'utilitzarà per enviar la cerca al sistema i començar el procés.

A més totes les pàgines vindran encapçalades pel logotip de l'Ajuntament de Terrassa, i un petit menú per poder accedir a les diferents opcions del portal.



Figura 15: Capçalera del sistema iSAC

3.6.2. Especificació de formats individuals de la interfície de pantalla

L'abast del projecte estableix que es realitza el motor de cerca, i per tant només existeixen tres (3) pantalles que interaccionen amb el projecte.



Figura 16: Pantalla d'inci

Una d'elles és la pantalla d'inici, on bàsicament hi haurà la caixa de text per tal de nosaltres introduir la nostra consulta.

L'altre serà la pantalla de resultats, on es mostrarà el llistat de resultats obtinguts per part del sistema. També hi haurà la possibilitat de tornar a realitzar la cerca.

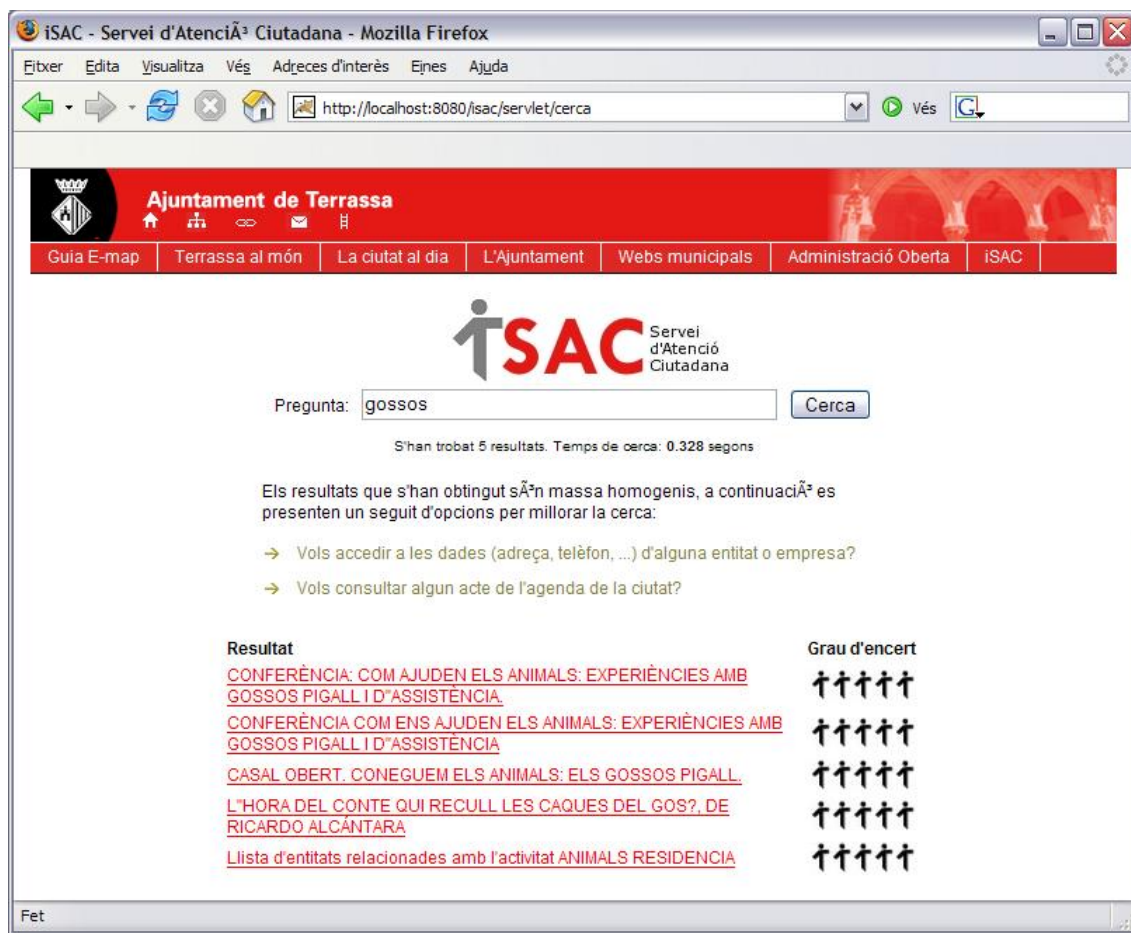


Figura 17: Pantalla de llistat de resultats

Finalment hi ha la pantalla on es mostra el resultat seleccionat per part de l'usuari o bé per part del sistema de forma directa. És el resultat que realment desitjàvem veure.

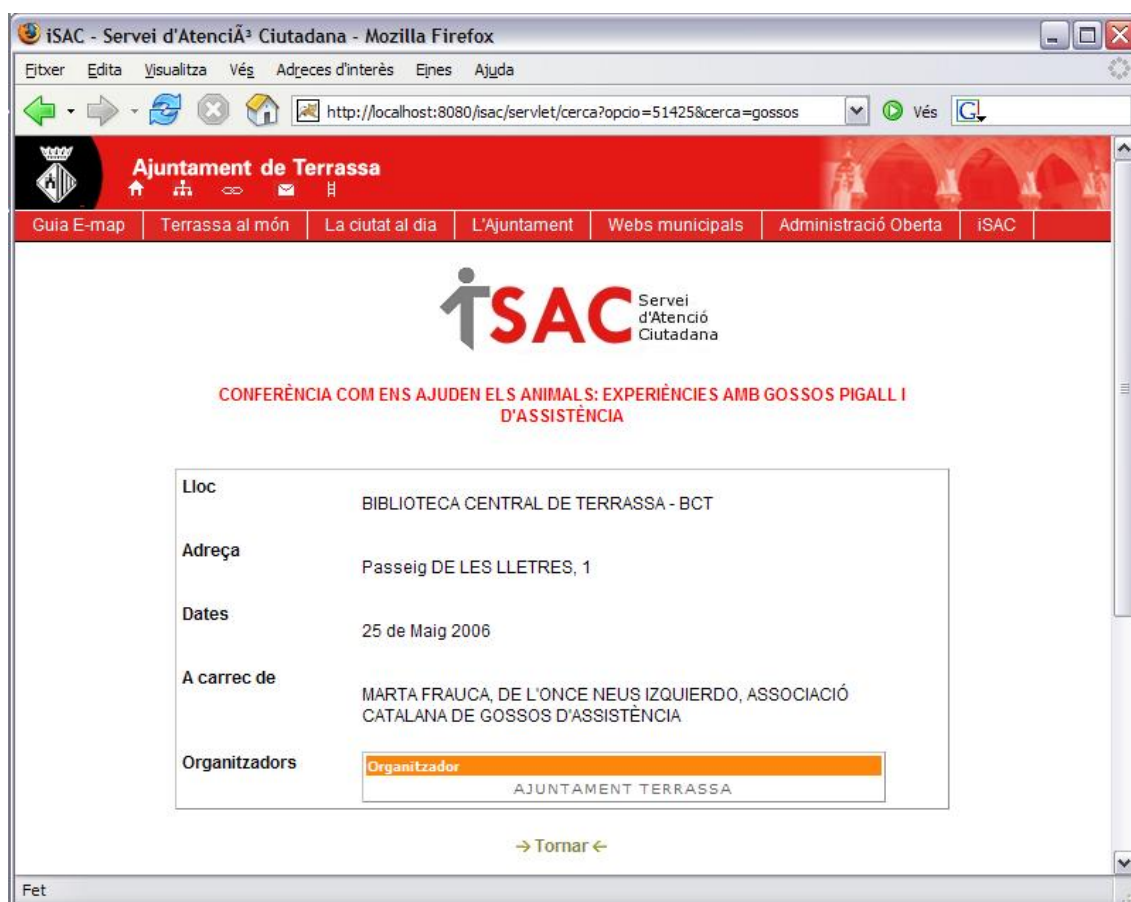


Figura 18: Pantalla amb el resultat

3.6.3. Identificació de perfils i diàlegs

El següent gràfic mostra la jerarquia que hi ha entre les diferents finestres o pantalles de l'aplicació.

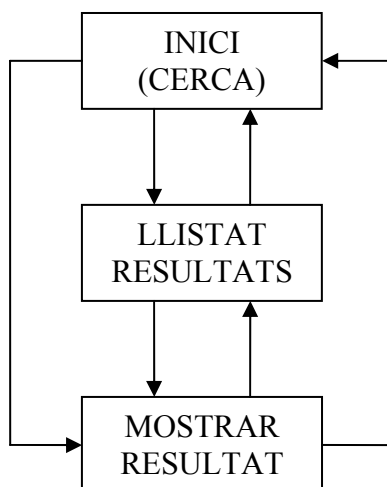


Figura 19: Diagrama de perfils i diàlegs

Capítol 4. Disseny del sistema d'informació (DSI)

En aquest capítol s'exposen les diferents activitats que contempla la Mètrica v.3 en el disseny del sistema d'informació.

Concretament, es detallaran les activitats següents:

- Definició de l'arquitectura del sistema
- Disseny físic de dades
- Generació d'especificacions de construcció

4.1. Definició de l'arquitectura del sistema

L'arquitectura del sistema iSAC es pot descompondre en tres parts:

- Una d'elles és el servidor de dades iSAC. Aquesta base de dades està formada per totes les entitats anomenades en els punts anteriors.
- El servidor iSAC també el podem subdividir en diferents fases:
 - El driver JDBC que serà l'encarregat d'interactuar amb la base de dades.
 - Els mòduls iSAC pròpiament dits, que seran els encarregats de realitzar la cerca i retornar els resultats.
 - El controlador de servlets.
 - HTTP Broker encarregat de passar del navegador cap al controlador de servlets.
- El navegador web serà el mètode utilitzat per l'usuari per realitzar les consultes al sistema.

A la següent figura podem veure l'estructura del sistema iSAC.

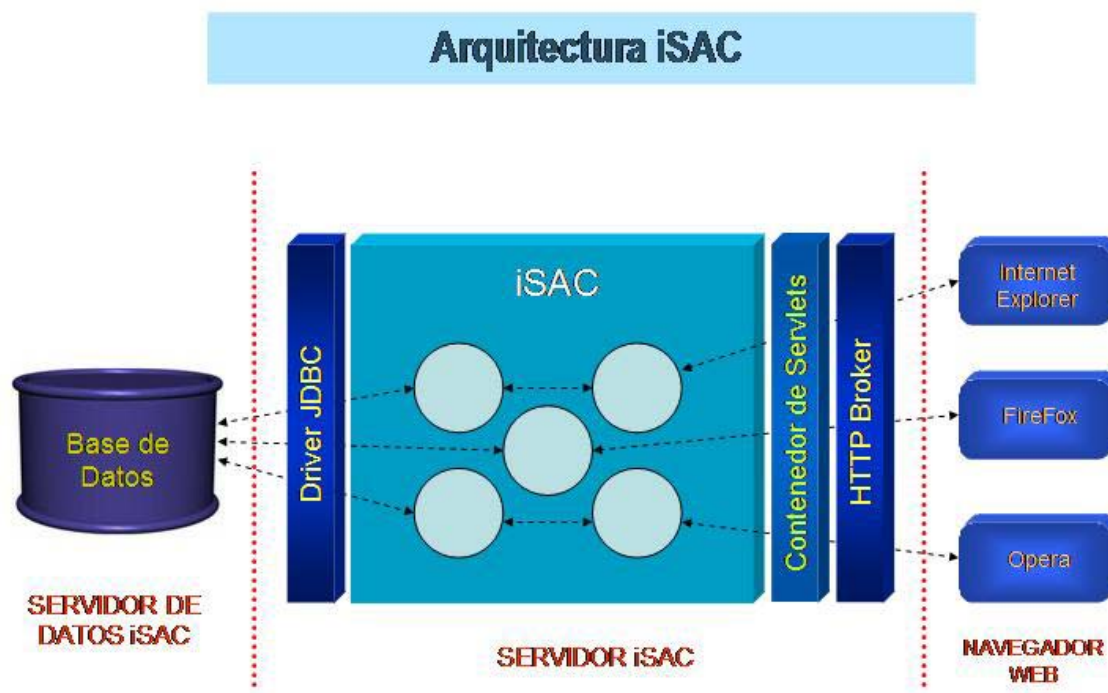


Figura 20: Arquitectura iSAC

A continuació desglossarem la unitat central, el servidor iSAC, per tal de poder observar com és de manera més detallada.

Aquesta part també el podem subdividir en tres blocs:

- Els mòduls complementaris, corresponents a les dades que existeixen actualment a l'administració pública. A partir d'aquestes dades n'extraurem les FAQs amb els seus descriptors i pesos. Aquesta serà la feina del Distiller.
- El nucli iSAC on podem destacar diferents parts:
 - L'accés a les dades de la base de dades.
 - La interfície d'usuari.
 - El suport per l'idioma, en aquest cas únicament català.
 - La gestió de les FAQs on es poden modificar els diferents diccionaris que conté el sistema.
 - El buscador de FAQs. És pròpiament el motor de cerca del sistema.
- A més es tenen uns altres mòduls on poden afegir idiomes addicionals.

Ara podem observar de manera més detallada el sistema iSAC i per tant observar aquests mòduls de l'arquitectura.

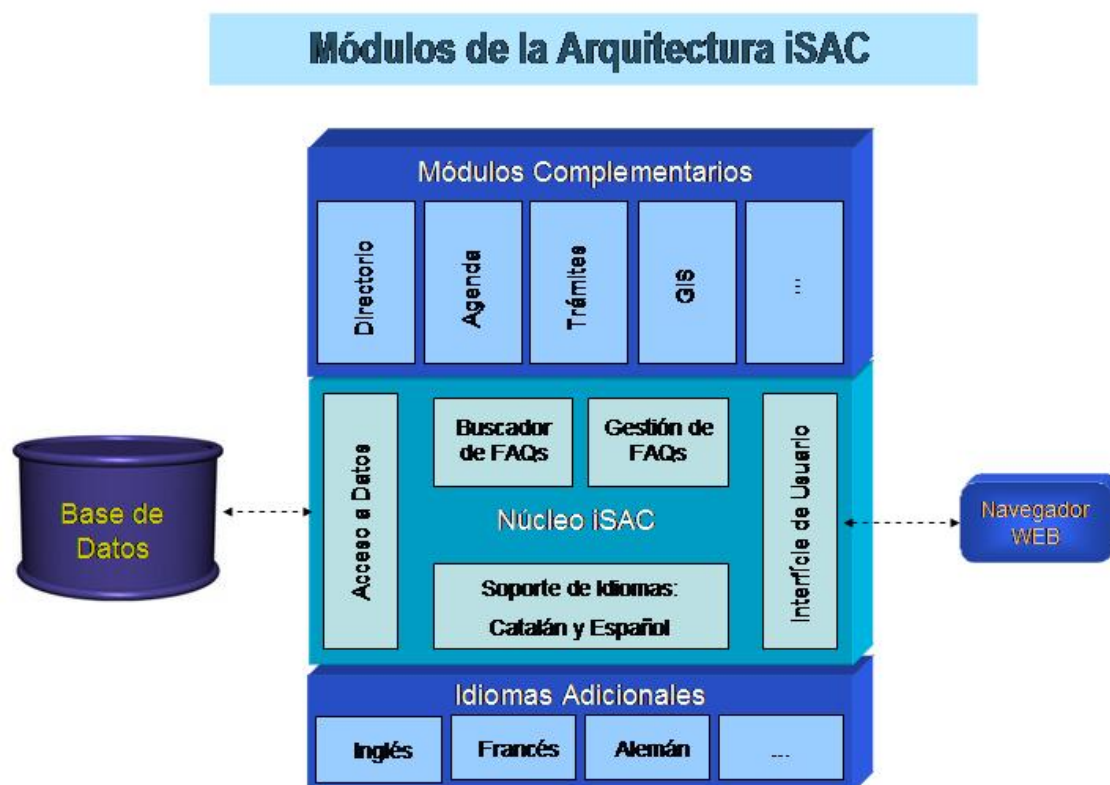


Figura 21: Mòduls de l'arquitectura iSAC

A la següent figura podem veure l'estructura iSAC, de manera molt detallada, per a un ajuntament gran amb el mòdul d'integració de bases de dades –Distiller– per extreure informació rellevant al sistema. Com ja hem comentat aquest mòdul no entra dins l'abast del projecte.

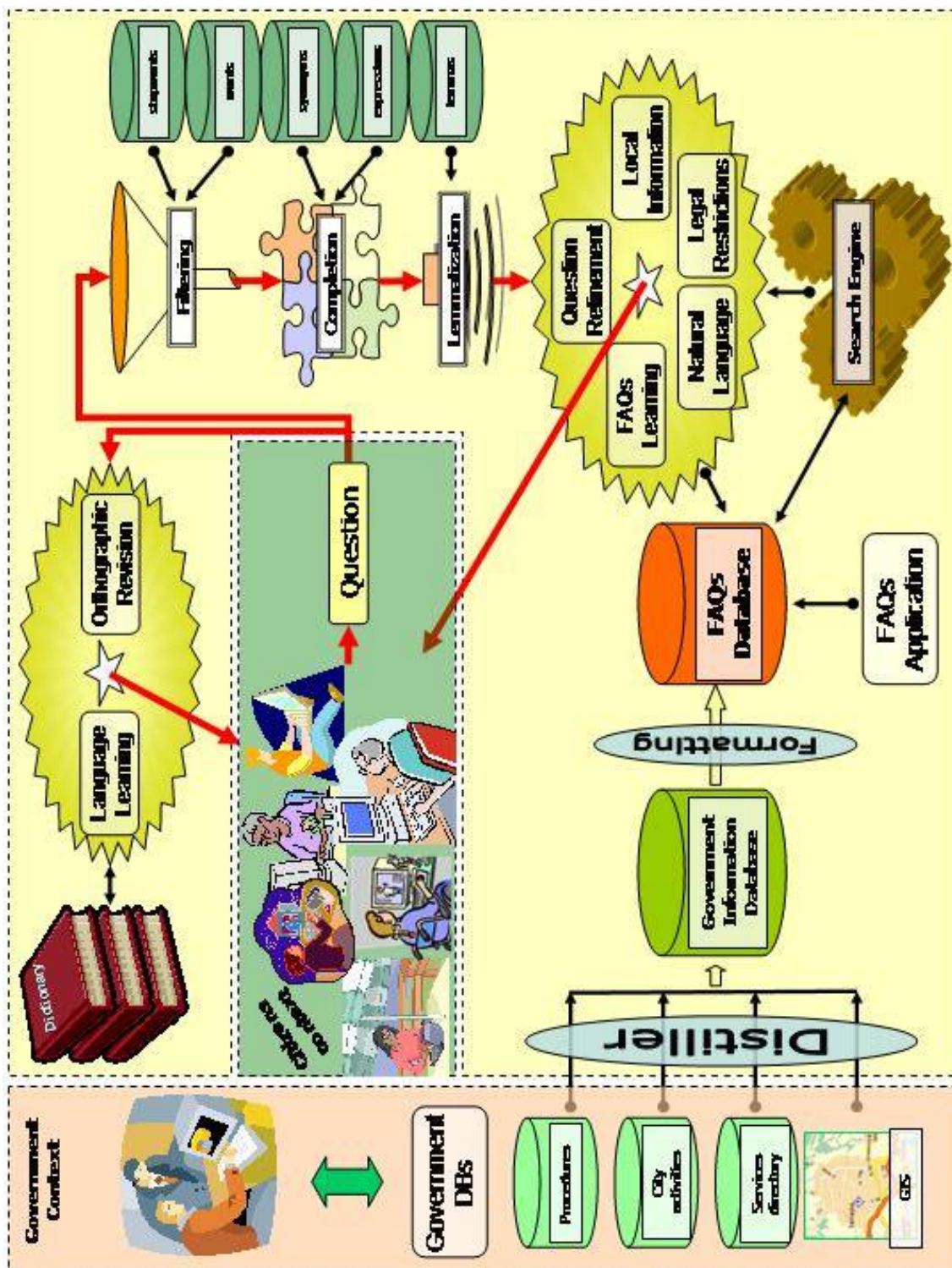


Figura 22: Arquitectura iSAC per a un ajuntament gran

Aquí pel contrari podem observar l'arquitectura per un ajuntament petit que no disposa de les dades d'informació ja estructurada i necessita utilitzar el sistema prototip per ordenar informació i posar-la a disposició del sistema automàtic d'atenció ciutadana.

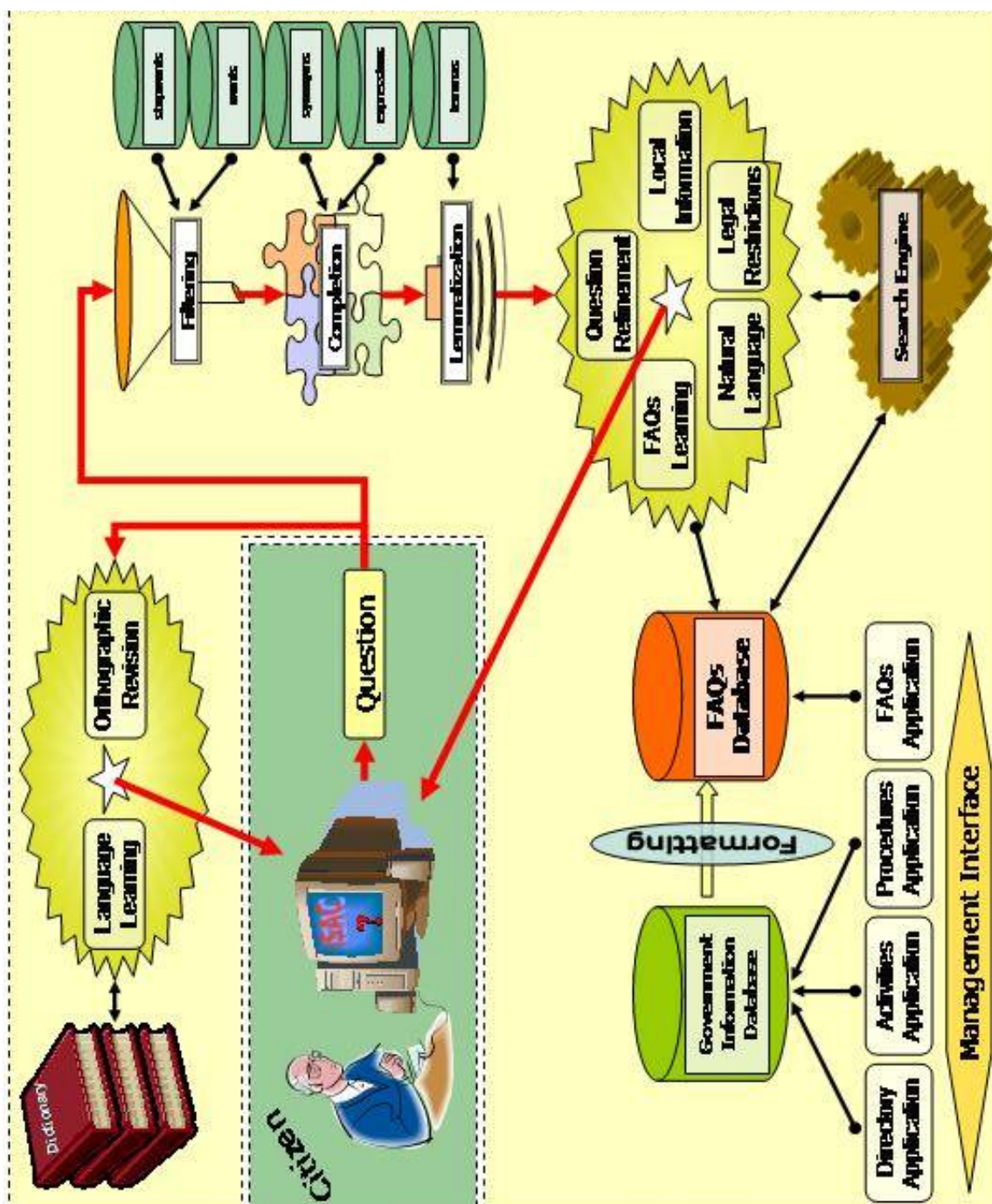


Figura 23: Arquitectura iSAC (versió extensa) ajuntament petit

4.2. Disseny de casos d'ús reals

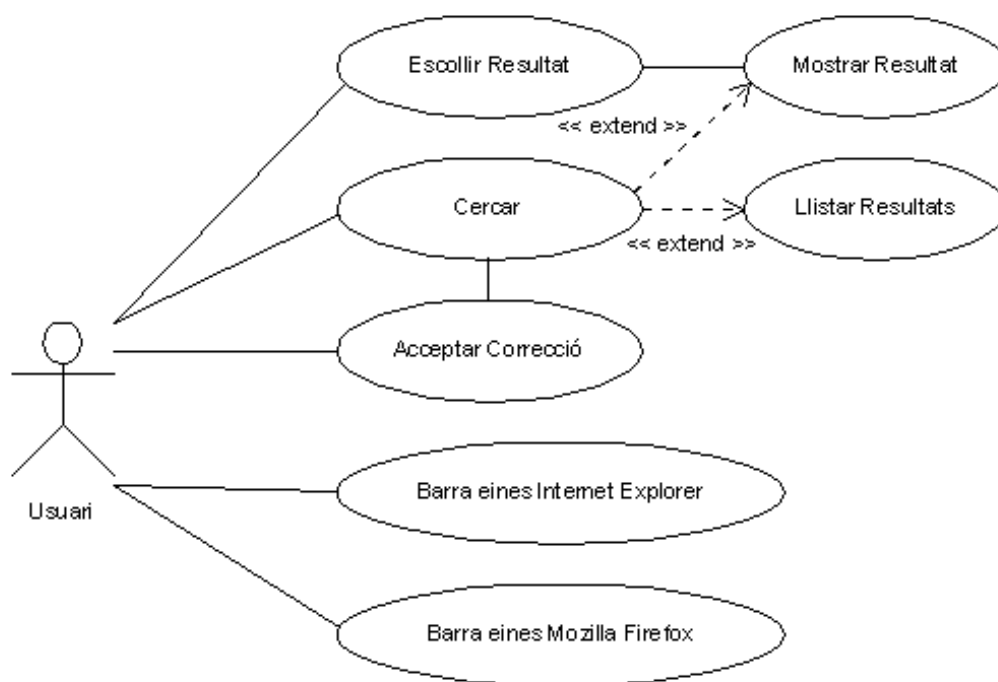


Figura 24: Diagrama de casos d'ús

4.3. Disseny de classes

En aquest apartat es realitza el disseny de les classes especificades a l'apartat de l'anàlisi de classes.

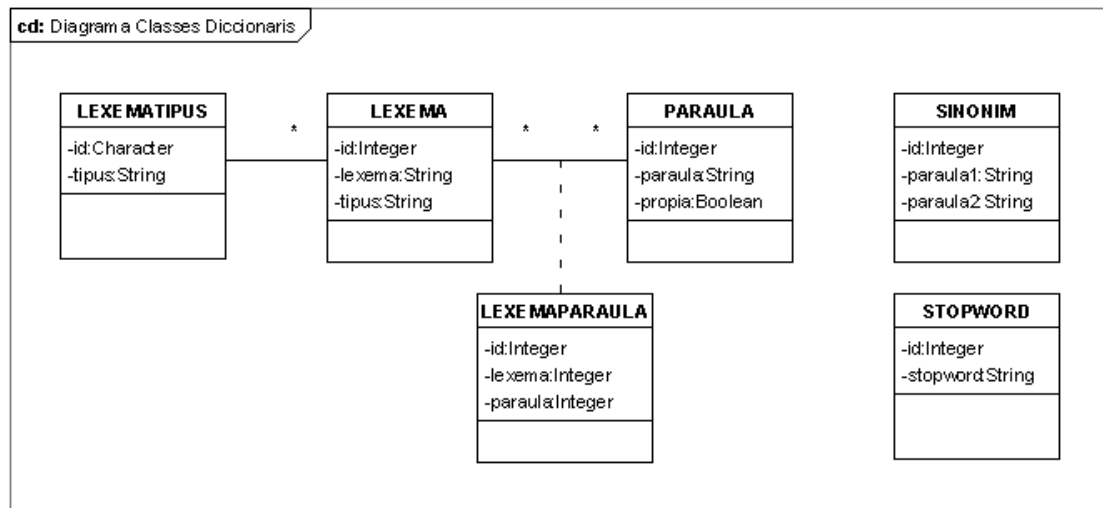


Figura 25: Diagrama de classes dels diccionaris

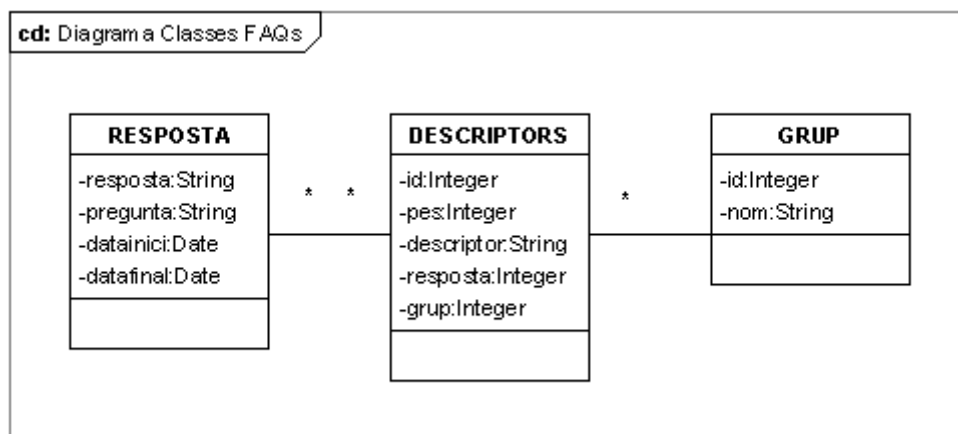


Figura 26: Diagrama de classes de les FAQs

4.4. Disseny físic de dades

En aquest apartat es realitza el disseny físic de dades del sistema, a partir dels models de classes de disseny del sistema.

4.4.1. Transformació del model de classes al model físic de dades

A continuació s'exposen les taules, columnes, claus primàries, claus secundàries (o foranes) i índexs resultants de l'aplicació de les regles de transformació.

LEXEMATIPUS:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	text	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
tipus	text	30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

LEXEMA:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
lexema	text	30	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
tipus	text	1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

PARAULA:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
paraula	text	50	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
propia	booleà		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

PARAULALEXEMA:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
lexema	enter		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
paraula	enter		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

SINONIM:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
paraula1	text	50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
paraula2	text	50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

STOPWORD:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
stopword	text	50	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

RESPOSTA:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
resposta	text	indefinit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pregunta	text	150	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
datainici	data		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
datafi	data		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pes	real		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

DESCRIPTORS:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
pes	real		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
descriptor	text	150	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
resposta	enter		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
grup	enter		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

GRUP:

<i>Atribut</i>	<i>Tipus</i>	<i>Tamany</i>	<i>CP</i>	<i>CF</i>	<i>Índex</i>
id	enter		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
grup	text	50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

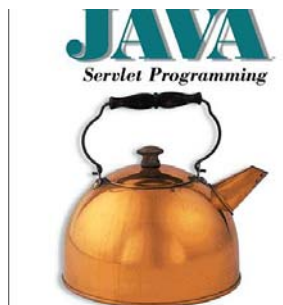
4.5. Especificacions de construcció

En aquest apartat es decideixen i s'analitzen les diferents eines i entorns per desenvolupar el projecte.

4.5.1. Especificacions de l'entorn de construcció**4.5.2. Llenguatge de programació**

Normalment el llenguatge de programació ve lligat pel servidor on pretenem allotjar l'aplicació. En aquest cas, Ajuntament de Terrassa, ens ha donat total llibertat per escollir tan el hardware com el software d'aquest servidor. En aquesta ocasió no ens hem trobat amb aquest impediment.

Descartat l'obstacle del servidor, s'ha optat per desenvolupar l'aplicació amb JSP. Aquest tipus de llenguatge inclou HTML juntament amb scripts de Java.



El fet de decidir aquest llenguatge ha vingut determinat pel fet que l'entorn Java és conegut i utilitzat arreu i per tant simplifica el fet d'haver de desenvolupar l'aplicació entre diferents integrants, així com facilita la feina a possibles millores i modificacions per part d'altres desenvolupadors.

Independentment d'aquest fet, el llenguatge JSP ja era un llenguatge conegut per mi, ja que durant la carrera l'he utilitzat en diverses ocasions.

4.5.3. Entorn de desenvolupament

El servidor on s'ha realitzat el desenvolupament i les proves d'execució té instal·lat un sistema operatiu Linux, més concretament una SuSE 9.0.



Aquest detall no és significatiu per l'aplicació, ja que funciona sota qualsevol sistema operatiu que tingui el servidor.

El fet d'utilitzar aquest en concret ha estat bàsicament pel fet de ser el servidor que tenia a l'abast. També considero que Linux és un sistema més estable i segur que d'altres. També s'ha tingut en compte el fet de ser un software de codi obert i en la gran majoria de casos de lliure distribució (no és el cas de SuSE) i per tant no s'ha de realitzar una despesa extra.

SuSE Linux és una de les més conegudes distribucions existents a nivell mundial. Entre les principals virtuts d'aquesta distribució es troba el fet que sigui una de les més senzilles d'instal·lar i administrar, ja que compta amb varis assistents gràfics, en especial, per la seva gran eina d'instal·lació i configuració YaST.

El servidor web utilitzat en la fase de proves és el Tomcat. Es tracta d'un servidor web amb suport de servlets y JSPs. Inclou el compilador Jasper, que compila JSPs convertint-les en servlets. El motor de servlets de Tomcat sovint es presenta en combinació amb el servidor web Apache.



The **Apache Jakarta Project**
<http://jakarta.apache.org/>

Ja que Tomcat va ser escrit en Java, funciona en qualsevol sistema operatiu que disposi de la màquina virtual.

Pel que fa al sistema gestor de bases de dades, hem utilitzat el PostgreSQL. És un servidor de base de dades relacional lliure, sota la llicència BSD. És una alternativa a altres sistemes de bases de dades de codi obert (com MySQL, Firebird y MaxDB), així com sistemes propietaris com Oracle o DB2.



PostgreSQL és un sistema de disseny visual que integra el disseny de la base de dades, el modelat, la creació i el manteniment. És molt senzill d'utilitzar. Entre altres coses, disposa d'un motor d'enginyeria inversa per recuperar un model de la base de dades i funcions de sincronització que apliquen els canvis del model automàticament a la base de dades. Els models es poden guardar en XML, de manera que és senzill extreure'n documentació. També disposa de constructors de consultes SQL.

4.5.4. Navegador

Es fa necessari també un navegador web per visualitzar a la banda client les pàgines que anem realitzant. En el nostre cas utilitzarem l'Internet Explorer versió v.6 i el Mozilla Firefox v.1.5.



Però per no restringir l'aplicació a aquests dos navegadors la provarem també en Netscape, Opera i Safari.



El que si ha de quedar clar, que el que no es pretén és desenvolupar un aplicatiu compatible amb les versions més antigues d'aquests navegadors, ja que suposaria un pas enrere amb l'avanç dels nous recursos que ofereix actualment la xarxa internet.

El fet, però, de basar-nos especialment en dos navegadors ha fet que incloguéssim a l'aplicació la possibilitat d'instal·lar una barra d'eines del nostre sistema iSAC per aquests navegadors. Aquesta possibilitat està inclosa dins la pantalla inicial del sistema.

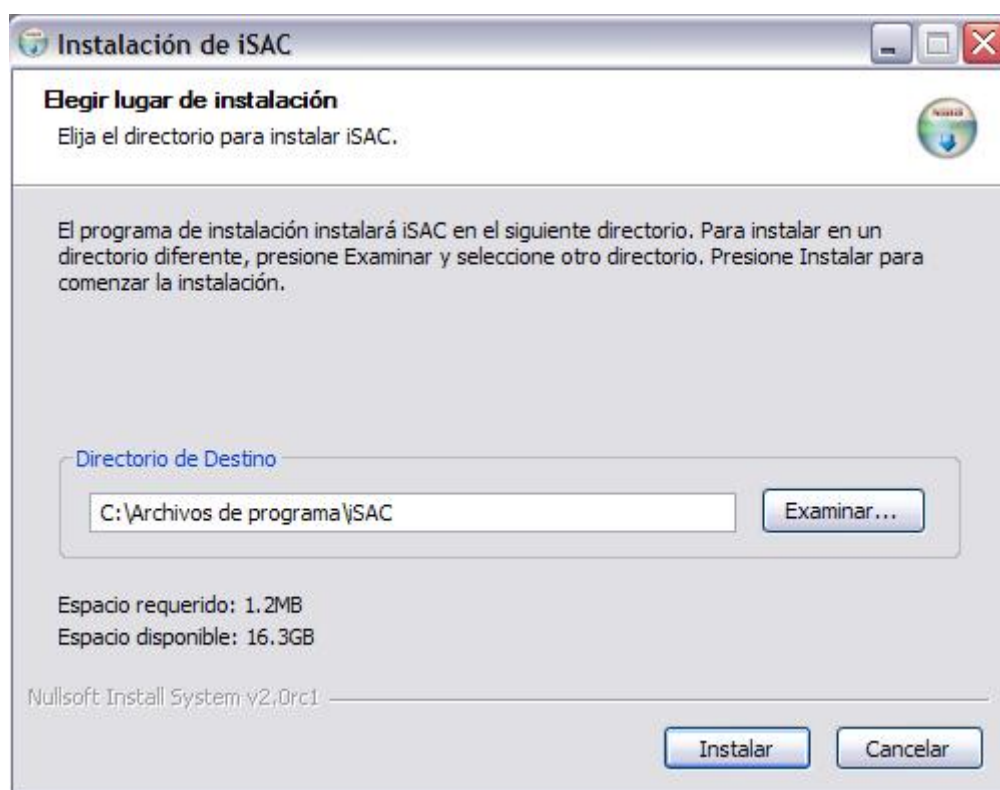


Figura 27: Barra d'eines iSAC

A continuació podem veure pas a pas el procés d'instal·lació d'aquestes eines per cadascun dels navegadors.

Instal·lació barra d'eines de Microsoft Internet Explorer

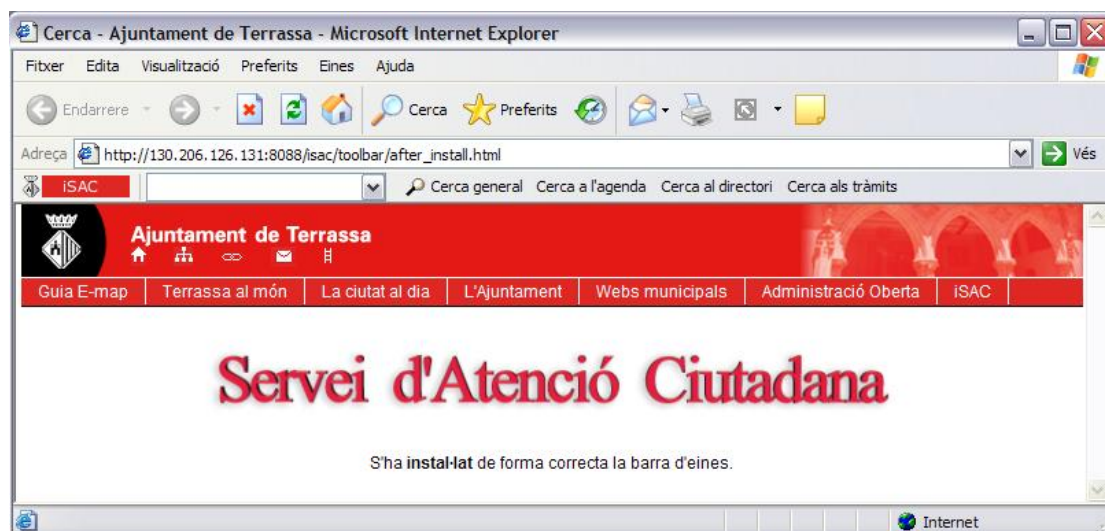
Després d'acceptar totes les peticions de seguretat que demana el Microsoft Windows, procedim a la instal·lació de la barra d'eines. Primer haurem d'escollir el directori allà on volem instal·lar l'aplicació.



Un cop seleccionat el directori, es procedeix amb la instal·lació.



Un cop finalitzada, automàticament s'obrirà de nou el navegador amb la nova barra d'eines.



Instal·lació barra d'eines de Mozilla Firefox

Primerament es descarrega la barra d'eines i se'ns demana la nostra confirmació per tal de procedir amb la instal·lació.



Un cop instal·lada de barra d'eines, es pot optar a desinstal·lar o bé cercar si hi ha alguna actualització disponible. Escollim aquesta segona opció.



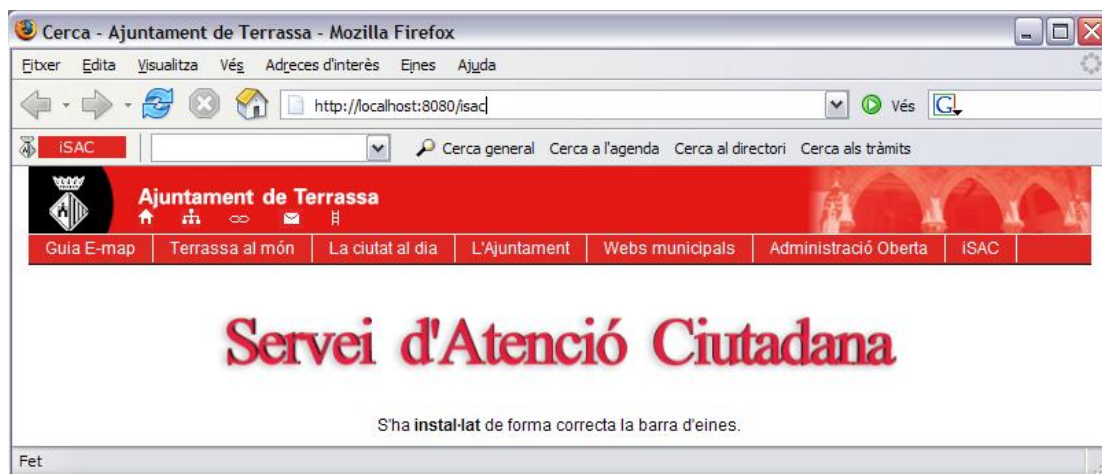
El sistema ens informa que actualment hi ha una actualització disponible i se'ns torna a sol·licitar el consentiment per tal d'instal·lar-la.



Finalitzada l'actualització, el sistema ens informa del final de la instal·lació.



Un cop reiniciat el navegador, podem veure com ja es disposa de la barra d'eines.



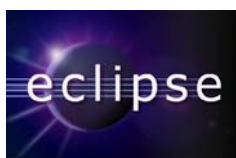
La barra d'eines manté els estàndards definits per totes les pantalles, essent una barra d'eines senzilla i sense peu a confusions.

4.5.5. Altres eines software

Les altres eines software utilitzades, al igual que ens els casos anteriors, segueixen la nostra política de cost 0: software gratuït, programari lliure.

Eines de disseny:

Per tal de realitzar les pàgines JSP, hem utilitzat Eclipse, ja que és un editor molt pràctic d'utilitzar.



Eclipse és una potent i completa plataforma de programació, desenvolupament i compilació d'elements tan variats com llocs web, programes en C++ o aplicacions Java.

Es tracta d'un entorn de desenvolupament integrat (IDE) en el que pots trobar totes les eines i funcions necessàries pel teu treball, recollides a més en una atractiva interfície que la fa més fàcil i agradable d'utilitzar.

Disposes amb un editor de text on pots veure el contingut del fitxer en el qual estàs treballant, una llista de feines, i d'altres mòduls similars. Si bé les funcions d'Eclipse són més aviat de caràcter general, les característiques del programa es poden ampliar i millorar mitjançant l'ús de plug-ins.

Transferència de fitxers al servidor:

Per a transferir els fitxers al servidor es pot fer servir tant FTP com FTPS (FTP sobre TLS). Nosaltres hem utilitzat el programa WinSCP.



Amb WinSCP pots connectar-te a servidors SSH utilitzant connexions segures com SCP (Secure Copy Protocol) o SFTP (Secure File Transfer Protocol), protegint així de forma total les dades transferides en dites connexions.

WinSCP resulta molt senzill d'utilitzar gràcies a la seva interfície, que pots utilitzar amb el format estàndard de l'Explorador de fitxers de Microsoft Windows o a l'estil Norton Commander, és a dir, amb dos finestres (una pel disc local i l'altre pel PC remot).

Podràs realitzar tot tipus d'operacions a la màquina remota gairebé com si es tractés del teu PC: comandes bàsiques de treball amb fitxers, utilitats de cerca i sobreescritura,... Pots utilitzar també un editor integrat i passar fitxers d'un PC a l'altre simplement arrossegant-los amb el ratolí.

WinSCP et garanteix la seguretat de les teves connexions i si vols, genera un log amb totes les activitats realitzades pel programa.

Capítol 5. Construcció del sistema d'informació (CSI)

En aquest capítol s'exposen les diferents activitats que contempla la Mètrica v.3 en la construcció del sistema d'informació.

Concretament, es detallaran les activitats següents:

- Preparació de l'entorn de generació i construcció.
- Execució de les proves
- Elaboració de manuals d'usuari
- Definició de la formació de l'usuari final

Degut a que l'abast del projecte no contempla la construcció del sistema d'informació, degut a que aquest abast és una petita part d'un gran projecte, es descriu a continuació, petits detalls de com es considera que s'hauria de realitzar aquest apartat de la Mètrica v.3.

5.1. Preparació de l'entorn de generació i construcció.

Com s'ha exposat a l'apartat 4.3.1 Especificacions de l'entorn de construcció el sistema gestor de base de dades emprat és PostgreSQL. Serà doncs amb aquesta aplicació que es generaran les taules, índexs, claus primàries i relacions.

S'instal·larà l'entorn de desenvolupament i el llenguatge de programació.

5.2. Execució de les proves

S'executen les proves unitàries per a cada component creat.

També cal executar les proves d'integració entre els diferents components. Es validen els resultats obtinguts.

En les proves del sistema es realitzen les comprovacions necessàries per tal de validar que el funcionament complert de l'aplicació és l'esperat, que no hi ha errors irreversibles i que respon davant fallades inesperades.

5.3. Elaboració dels manuals d'usuari

S'elaborarà un manual d'usuari en el que es definirà el funcionament de l'aplicació, definint totes les interfícies i opcions de treball. Amb aquest document es pretén donar a l'usuari (en aquest cas Ajuntament de Terrassa) un suport per la resolució dels dubtes que puguin sorgir durant la seva utilització.

El contingut que ha de compondre aquest manual és el següent:

- Explicació del funcionament de l'aplicació a partir dels seus menús i pantalles, comentant tots els diàlegs que hi apareixen.
- Exemples pràctics sobre el funcionament i la gestió de l'aplicació.

5.4. Definició de la formació de l'usuari final

Com s'ha comentat, l'usuari de l'aplicació (de gestionar-lo i encarregar-se del seu manteniment) és l'Ajuntament de Terrassa. Les persones encarregades del seu manteniment i funcionament són persones que han indicat els requeriments de l'aplicació i que per tant els hi serà molt senzill l'adaptació amb l'aplicació, ja que s'ha intentat en tot moment que fos el màxim visual i intuïtiva possible.

La formació contemplarà els següents aspectes:

- Explicació detallada de les funcionalitats comuns de l'aplicació.
- Se li detallarà la funcionalitat de cada una de les opcions disponibles i les jerarquies de les pantalles del sistema d'informació.
- Funcionament dels manteniments de dades.
- Finalment, se li remarcarà la importància de la realització de còpies de seguretat periòdiques.

Un cop fet tot això, es passarà a la utilització pràctica de l'aplicació amb la resolució dels dubtes que apareguin sobre el seu funcionament, indicant quina és la utilització òptima per treure'n el millor rendiment.

Part III

Ampliacions i millores

En aquesta part s'expliquen aquelles ampliacions i millores que es considera que ha de tenir el sistema per tal d'avançar.

Capítol 1. Ampliacions i millores

Està clar que en un projecte com aquest, les ampliacions i millores poden ser múltiples. Ara bé, centrant-nos bàsicament en el motor de cerca, podem dir que les ampliacions més importants podrien ser:

- Expansió multilingüe (5 idiomes: català, castellà, anglès, àrab i francès). El perfil de la ciutadania és actualment molt canviant, no sols en quantitat i en composició social, en nivell d'exigència, també en qualitats culturals i humanes, per la integració de gran diversitat de col·lectius de procedència molt diferent i distant geogràficament parlant. La majoria d'aquesta nova demanda no parla correctament l'idioma habitual de relació dels actuals serveis (català i castellà). Així mateix, tampoc són turistes, que poden ser atesos en serveis especialitzats ja preparats amb capacitats multilingües. Algunes vegades, que no la majoria, a més, es tracta de col·lectius amb necessitats especials pel seu nivell educatiu realment baix.

L'extensió multilingüatge serà paulatina i de manera natural. Ampliar a idiomes de la mateixa arrel (la llatina) no tindrà les mateixes dificultats de reconeixement que quant ens proposem abordar una altra arrel (l'àrab o l'anglès per exemple). En tot cas serà l'evolució de determinades circumstàncies socials, polítiques o comercials que determinarà afegir esforços en aquest sentit.

D'altres millores més generals que ajudarien a completar l'equipament dels serveis d'Informació i Atenció a la Ciutadania, permetent la continuïtat en el desplegament de nova tecnologia de relació ciutadans – administració amb la incorporació de noves eines de comunicació, d'intel·ligència artificial, de traducció de missatges relacionals, etcètera, podrien ser:

- Que l'usuari pugui fer valoracions del sistema i per tant ajudar a la millora de l'aplicació.
- A partir de la navegació de l'usuari guardar logs de tots els procediments que fa. A partir d'aquests logs, poder carregar les dades que després ens siguin útils a una base de dades del servidor. A partir d'aquestes dades emmagatzemades, poder consultar i saber com utilitza l'usuari el sistema.
- Obtenció de resultats de consultes. Poder observar l'evolució del sistema i com es comporta. L'administrador podrà visualitzar resultats com ara quines són les cerques més buscades, quines paraules utilitza més la gent per realitzar la cerca, quin és el temps de resposta del sistema (en franges horàries per tal de poder prevenir problemes d'eficiència), poder visualitzar les valoracions de les FAQ's i del sistema,...

- Facilitació de la veu i la conversa (a l'oficina i a casa, per telèfon fix, al carrer, via mòbil i SMS). Les modernes tecnologies de reconeixement de veu ja permeten l'accés amb garanties per via telefònica als sistemes d'informació automatitzada. La incorporació d'aquesta tecnologia a la nova estructura i capacitats del servei possibilitarà l'accés 24 hores i sense necessitat d'intervenció directa de personal per resoldre les demandes més freqüents de la ciutadania des de qualsevol lloc i a qualsevol hora que es produeixi la necessitat. A més, potenciarà l'ús de les tecnologies de reconeixement de veu per a l'atenció de demandes amb comprensió en una sola intervenció de les persones usuàries, sense necessitat de navegar per un llarg i tediós sistema d'arbre per concretar la resposta, fruit de la combinació de la tecnologia de veu amb el llenguatge natural.
- Adequació a la multicanalitat (a casa amb televisió digital-TDT, e-mail,...). El desenvolupament dels nous formats de la televisió digital terrestre (TDT) ja són en fase prou avançada per començar a preveure'n un ús important en les relacions de l'administració amb la ciutadania en molts casos d'incapacitat o mobilitat impedida temporalment o permanent, persones que viuen soles i necessiten cura a domicili,... S'estan fent avanços importants en aquests serveis, però la posta a punt de la tecnologia requerirà un important cos de recerca en productes, prèvia a la sortida al mercat, i que les empreses difícilment podran assumir.

<i>FITES</i>	<i>BENEFICIS</i>
Ampliació del reconeixement idiomàtic a altres llengües	Afavorir la integració cultural
	Afavorir la igualtat de tracte
Aplicació de metodologia d'ESCALA DE VALORS, a través d'agents intel·ligents i d'eines d'intel·ligència artificial a tasques de back-office	Incorporació del feedback ciutadà a la millora de gestió i al redisseny de les polítiques públiques
	Millora de l'objectivitat dels resultats, evitant els biaixos conscients o involuntaris de l'enquesta
Ampliació dels canals d'accés	Noves aplicacions mòbils
	Millora de la interactivitat de les aplicacions web en serveis en xarxa
	Reconeixement de veu en atenció telefònica

Part IV

Conclusions

En aquesta part s'exposen quines han estat les conclusions extretes de la realització del projecte.

Capítol 1. Conclusions

Aquest projecte final de carrera va arribar a mi després d'estar un any treballant com a becari en un SpinOff de la Universitat de Girona, Agents Inspired Technologies S.A. El cap de projectes de l'empresa, Miquel Montaner (que a la fi ha estat el meu director), juntament amb el que és el meu tutor Josep Lluís de la Rosa, em van donar l'oportunitat de realitzar aquest projecte de desenvolupament com a projecte final de carrera.

El fet de compaginar estudis i treball, em suposava un esforç molt gran realitzar un projecte final de carrera, i aquesta oportunitat em va permetre aprofitar les hores de feina per tal de realitzar-lo.

Ha estat un repte molt important per mi haver participat en moltes de les tasques que comprèn un projecte d'aquest abast: entrevistes amb el client i entre l'equip de treball, sintetització d'idees, confecció dels requeriments, fer l'anàlisi i el disseny de parts de l'aplicació, programar-la, validar-ne les funcionalitats,...

Un punt important en el decurs del projecte ha estat el tracte amb el client. En aquest aspecte he viscut d'una forma més propera tot els aspectes relacionats amb les relacions humanes i aspectes de caire empresarial. D'aquest fet n'he tret la conclusió que la funcionalitat de les aplicacions és el més important, i no la forma com s'han realitzat.

Analitzant pròpiament el projecte, puc dir que un cop finalitzat el projecte cal valorar tots els punts marcats en els objectius. S'han obtingut satisfactòriament els objectius marcats i s'ha comprovat que es compleixen tots els requisits demanats: s'ha desenvolupat el motor de cerca d'un sistema al qual es poden adreçar tot els ciutadans, entenent el llenguatge col·loquial, essent capaç de preveure errors ortogràfics, millorar la cerca afegint sinònims, obtenint el lema de les paraules per tal d'ampliar la cerca i interpretant expressions populars, regionalismes i barbarismes. El sistema és capaç de respondre a la consulta del ciutadà, tot i que s'expressi d'una manera diferent a la que faríem qualsevol de nosaltres.

S'han complert objectius d'estudi i aprenentatge com ara, l'ús d'una metodologia com la Mètrica en la seva versió 3 amb la que s'ha realitzat un estudi estructurat i s'han seguit unes pautes ben definides.

S'han posat en pràctica coneixements adquirits al llarg de la carrera sobre mètriques, llenguatges de programació, anàlisi i disseny de bases de dades, enginyeria del software i alternatives en el disseny d'aplicacions.

Tanmateix, he aprofitat l'ocasió per involucrar-me en l'aprenentatge dels *servlets de Java*, que m'han permès conèixer noves alternatives en la construcció de sistemes d'informació.

L'abast d'aquest projecte final de carrera ja s'ha assolit, tot i que el projecte que l'engloba ha fet poc més que començar. Espero en aquests nous passos, seguir aprenent com ho he fet fins ara.

Part V

Documentació de referència

En aquesta part es mostra la documentació de referència consultada per la confecció de la documentació i del projecte.

Capítol 1. Bibliografia

Burke, R. D., Hammond, K. J. Kulyukin, V. A., Lytinen, S. L., Tomuro, N. and Schoenberg, S. Question answering from frequently asked question files: Experiences with the faq finder system. Technical report, 1997.

Burke, R., Hammond, K., Kulyukin, V., Lytinen, S., Tomuro, N., and Schoenberg, S. (1997). Natural Language Processing in the FAQ Finder System: Results and Prospects. In Papers from the 1997 AAAI Spring Symposium on Natural Language Processing for the World Wide Web., Stanford University, California.

González, G., Angulo, C. López, B. and de la Rosa J.LL. Smart User Models for Ambient Recommender Systems. In Proceedings of International Workshop on Ambient Intelligence and (Everyday) Life. pp.113-122. July 21 - 22, 2005. ISBN: 84-609-6632-1. San Sebastian, España.

González, G., López, B. and de la Rosa, J. LL. A Multi-agent Smart User Model for Cross-domain Recommender Systems. Proceedings of Beyond Personalization 2005: The Next Stage of Recommender Systems Research. International Conference on Intelligent User Interfaces IUI'05. January 9, 2005, San Diego, California, USA.

Guzmán, J., González, G., de la Rosa J.LL. and Castán, J.A. Modelling the Human Values Scale in Recommender Systems: A First Approach. Artificial Intelligence Research and Development. Vol. 131. López, B. Meléndez, J., Radeva, P. and Vitriá, J. (Eds.). IOS Press.405-412. Octubre 26-28, 2005. ISSN: 0922-6389. Alghero, Sardinia, Italy.

Jeon, J., W. Croft, B., and Lee., J. H. Finding similar questions in large question and answer archives. In Proceedings of the 14th ACM international Conference on information and Knowledge Management (Bremen, Germany, October 31 - November 05, 2005). CIKM '05. ACM Press, New York, NY, 84-90.

Metzler, D. and Croft, W.B. Analysis of Statistical Question Classification for Fact-based Questions, in Information Retrieval, 8(3), 481-504, 2005

Moldovan, D., Paşca, M., Harabagiu, S., and Surdeanu, M. 2003. Performance issues and error analysis in an open-domain question answering system. ACM Trans. Inf. Syst. 21, 2 (Apr. 2003), 133-154. DOI=<http://doi.acm.org/10.1145/763693.763694>

Montaner, M., López, B., de la Rosa, J.Ll., "Evaluation Of Recommender Systems Through Simulated Users". 6th International Conference on Enterprise Information Systems (ICEIS'04). Porto (Portugal). April, 2004.

Capítol 2. Referències web

<http://www.map.es/csi/metrica3> : Ministerio para las Administraciones Públicas. Metodología de planificación y desarrollo de sistemas de información.

<http://www.softonic.com> : descàrrega d'aplicacions i descripcions.

<http://httpd.apache.org> : pàgina principal apache.

<http://www.postgresql.org/> : pàgina web del SGBD PostgreSQL.

<http://www.java.com> : pàgina principal de java.

<http://sunsite.unam.mx/java.html> : manual de java.

<http://www.mundotutoriales.com> : tutorial de servlets de java.

<http://www.programacion.com> : exemples i petites aplicacions en java i applets de java.

<http://www.wikipedia.org> : enciclopèdia lliure en diversos idiomes.

<http://lingucomponent.openoffice.org> : pàgina web del diccionari OpenOffice.

<http://garraf.epsevg.upc.es/freeling> : diccionari pels lexemes.

<http://www.google.com> : buscador web.